مجله مدل‌سازی پیشرفته ریاضی
سال ۱۴۰۱، دوره ۱۲، شماره ۴، ص ۴۹۴-۵۰۵

شاپا چاپی: ۲۲۵۱-۸۰۸۸
DOI: 10.22055/JAMM.2022.41081.2049

راهبردهای وزنی جدید برای نمونه‌گیری فضایی همبسته پواسون

هادی فرخی نیا ×، رحیم چینی پرداز، غلامعلی پرهام

گروه آمار، دانشکده علوم ریاضی و کامپیوتر، دانشگاه شهید چمران اهواز

تدبیر مسئول: غلام‌رضا محتشمی برزادران

تاریخ دریافت:۱۴۰۱/۳/۲۳ تاریخ پذیرش: ۱۴۰۱/۹/۱۵

چکیده:
در نمونه‌گیری پواسون هر واحد مستقل از دیگر واحدها با احتمال شمول معین انتخاب می‌شود و اندازه‌ی نمونه $n(s)$ متغیری تصادفی با نوسان زیاد است. همچنین اگر بین واحدهای جامعه همبستگی یا روندی وجود داشته باشد، کارایی برآورد پارامتر جامعه کاهش می‌یابد. نمونه‌گیری همبسته‌ی پواسون (CPS[۱]) و نمونه‌گیری فضایی همبسته‌ی پواسون یا ($SCPS$[۲]) که تعدیلی از نمونه‌گیری همبسته‌ی پواسون است، روش‌های جایگزین معرفی شده برای نمونه‌گیری پواسون هستند. دراین روش‌ها با استفاده از راهبردهای وزنی، تغییرات اندازه‌ی نمونه کاهش می‌یابد و نمونه حاصل پراکندگی بیشتری روی‌کل جامعه دارد. هدف معرفی راهبردهای وزنی جدیدی است، که بوسیله شبیه‌سازی نشان خواهیم داد، نسبت به راهبردهای پیشین، کارایی برآورد پارامتر جامعه افزایش و تغییرات اندازه‌ی نمونه را کاهش می‌دهند.

واژه‌های کلیدی: برآوردگر هورویتز–تامپسون، نمونه‌گیری پواسون، نمونه‌گیری فضایی، نمونه‌گیری همبسته پواسون.

رده‌بندی ریاضی: 62D05; 62F07; 62F10

۱ مقدمه

نمونه‌گیری پواسون نوعی از روش نمونه‌گیری دنباله‌ای است، که در آن محقق واحدهای جامعه را پی در پی و به ترتیب مشاهده می‌کند و باید در زمان مشاهده واحد در مورد ورود به نمونه یا صرف نظر کردن از آن تصمیم بگیرد. این طرح نمونه‌گیری در جامعه متناهی با احتمال شمول نابرابر توسط هاژک [۶] ، گوس و کلارک [۹] سرندال و همکاران [۱۰] مورد مطالعه قرار گرفته است. در این روش برای هر عضو i از جامعه مورد مطالعه یک آزمایش برنولی مستقل با احتمال p_i انجام می‌شود، که برآمدهای آزمایش تصادفی برای واحد i به صورت تابع شاخص شمول I_i نشان داده می‌شود. بنابراین اگر واحد i ام وارد نمونه شود تابع شاخص شمول برابر با یک در غیر این‌صورت صفر است. در نتیجه اندازه‌ی نمونه $n(s) = \sum_{i=1}^{N} I_j$ یک متغیر تصادفی است، این ویژگی یعنی ثابت نبودن اندازه نمونه باعث کم شدن کارایی این

×نویسنده مسئول مقاله
رایانامه: hadi-farokhnia@stu.scu.ac.ir

روش در مقایسه با دیگر روش‌ها می‌شود، که در مثال‌های شبیه‌سازی شده نشان داده خواهد شد. همچنین اگر بین واحدهای جامعه همبستگی یا روندی وجود داشته باشد، با روش نمونه‌گیری پواسون نمی‌توان آن را در نظر گرفت و ممکن است، باعث اریبی برآورد پارامتر جامعه شود. از این رو میستر [۸] جامعه‌ی با حجم متناهی N را بررسی کرد، با فرض این که بین واحدهای مجاور همبستگی مثبتی وجود داشته باشد، به طوری که با انتخاب واحدهای مجاور اطلاعات یکسانی مشاهده می‌شود. این وضعیت معادل زمانی است که فقط یکی از واحدها انتخاب شود، در نتیجه انتخاب واحدهای مجاور ناکارآمد خواهد بود. بنابراین باندسون و توربون [۲] روش جایگزینی به نام نمونه‌گیری همبسته پواسون برای حل این مشکل معرفی کردند. در این روش با اختصاص وزن‌های نامنفی به شاخص‌های شمول، همبستگی منفی بین شاخص‌های شمول واحدها ایجاد می‌شود، که باعث کاهش تغییرات اندازه‌ی نمونه و در بعضی موارد صفر شدن تغییرات اندازه‌ی نمونه می‌شود. بنابراین نمونه‌ای بدست می‌آید ،که پراکندگی نمونه روی کل جامعه بیشتر خواهد شد. گرافستروم [۵] راهبردهای وزنی متفاوتی را برای نمونه‌گیری همبسته پواسون معرفی و مقایسه کرده است و روش نمونه‌گیری فضایی همبسته پواسون را گرافستروم [۴] در ادامه برای نمونه‌گیری با احتمال نابرابر در جوامعی که به طور فضایی در یک یا بیش از یک بعد گسترده شده است، ارائه داده است. التیری و همکاران [۱] ،یک سیستم وزن دهی که شامل آنتروپی فضایی متغیر مورد مطالعه است، را بررسی کردند. همچنین گرافستروم [۳] چندین روش مکان محوری را برای انتخاب نمونه‌های متعادل فضایی از جوامع فضایی معرفی کرده است. کاربردهای فراوانی از نمونه‌گیری فضایی را در مطالعات زیست محیطی می‌توان یافت. جامعه در مطالعات زیست محیطی اغلب روی فضا توزیع شده است، مانند پوشش گیاهی یک جنگل ،که اگر داده‌های سنجش از راه دور هم در دسترس باشند، می‌توان چنین اطلاعاتی را وارد طرح نمونه‌گیری کرد، تا مجموعه داده‌ها کامل‌تر باشد. روش نمونه‌گیری فضایی همبسته‌ی پواسون مجموعه‌ای از راهبردهای انتخاب وزن و شبیه نمونه‌گیری همبسته‌ی پواسون می‌باشد، با این تفاوت که باید فاصله‌ی بین واحدها نیز معلوم باشد.

در نمونه‌گیری از یک جمعیت متناهی به طور معمول، هدف برآورد $Y_T = \sum_{i=۱}^{n} y_i$ یعنی مجموع y_i ها است، که y_i مقدار صفت در واحد i ام می‌باشد. همچنین برای استفاده از اطلاعات کمکی، که در اغلب موارد مقادیر معلوم یک متغیر هستند و با x_i برای هر $i \in \mathcal{U} = \{i = ۱, \ldots, N\}$ نمایش داده می‌شود، هر واحد i را با احتمال شمول $p_i = cx_i$ که c یک مقدار ثابت و مثبت است، انتخاب می‌شود. برآوردگری که برای نمونه‌گیری با احتمال متغیر استفاده می‌شود، برآوردگر نااریب هورویتز –تامپسون [۷] است، که به صورت زیر محاسبه می‌شود:

$$\hat{Y}_{HT} = \sum_{i=۱}^{N} \frac{y_i}{p_i} I_i \qquad (۱.۱)$$

که I_i تابع شاخص شمول است. شکل نمونه‌ای برآوردگر هورویتز – تامپسون را می‌توان به صورت زیر تعریف کرد:

$$\hat{Y}_{HT}(S) = \sum_{i \in S} \frac{y_i}{p_i}.$$

به طوری‌که فقط مقدار نسبت $\frac{y_i}{p_i}$ مربوط به آن واحدهای که در نمونه S قرار می‌گیرند، محاسبه می‌شود.

راهبردهای وزنی که درمقالات مختلف تاکنون معرفی شده‌اند، همیشه دارای کارایی بهتر نسبت به وزن‌های صفر یا نمونه‌گیری پواسون نیستند. در این مقاله هدف معرفی راهبردهای وزنی جدیدی است، که عملکرد بهتری نسبت به راهبردهای پیشین دارند. بدین منظور این مقاله در پنج بخش آماده شده است، که در بخش دوم به نمونه‌گیری همبسته پواسون و معرفی راهبردهای وزنی جدید می‌پردازیم. روش نمونه‌گیری فضایی همبسته پواسون را در بخش سوم معرفی و بررسی می‌کنیم. بخش چهارم به مثال‌های شبیه‌سازی برای مقایسه روش‌ها و راهبردهای وزنی اختصاص داده شده است. در پایان نتیجه‌گیری و مباحثی که می‌توان در آینده مورد بررسی قرارگیرد، را در بخش پنجم ارائه می‌دهیم.

۲ نمونه‌گیری همبسته پواسون

جامعه‌ای را فرض کنید، که واحدهای آن یک به یک در زمان‌های $t_۱$ تا t_N مشاهده و به ترتیب از ۱ تا N شماره‌گذاری می‌شوند. در نمونه‌گیری همبسته پواسون باید در لحظه مشاهده کردن هر واحدi ، برای وارد شدن آن واحد با احتمال معلوم p_i در نمونه تصمیم‌گیری شود. البته ممکن است این احتمال به دلیل متناسب بودن متغیر کمکی، در هنگام مشاهده شدن واحد معلوم شود. بنابراین نمونه‌گیری همبسته پواسون با واحد اول که دارای احتمال شمول $p_۱$ است، شروع می‌شود. اگر واحد نخست وارد نمونه شود، تابع مشخصه شمول مربوط به آن واحد یعنی $I_۱$ برابر یک و در غیر این صورت برابر صفر خواهد شد. سپس احتمال شمول برای باقی واحدها به صورت زیر اصلاح یا به‌روز می‌شود :

$$p_i^{(۱)} = p_i - (I_۱ - \pi_۱) w_{i-۱}^{(۱)}, \qquad i = ۲, \ldots, N$$

در اینجا $w_{i-1}^{(1)},\ \ i=2,\dots,N.$ وزن مربوط به واحد i ام است. بنابراین دومین واحد جمعیت با احتمال جدید $p_2^{(1)}$ وارد نمونه می‌شود. پس از مشاهده دومین واحد و ثبت I_2 دوباره احتمال شمول برای باقی واحدها با استفاده از فرمول زیر به روز رسانی می‌شوند.

$$p_i^{(2)}=p_i^{(1)}-(I_2-p_2^{(1)})w_{i-2}^{(2)},\qquad i=3,\dots,N.$$

که در آن $w_{i-2}^{(2)},$ وزن‌های مربوط به هر واحد است، که می‌تواند متفاوت با وزن‌های قبلی باشد. این روند با مشاهده‌ی هر واحد تکرار می‌شود. به طور کلی در گام j ام ، وقتی واحد $j-1$ مشاهده شده و مقدار I_{j-1} ثبت شده باشد ، احتمال شمول برای باقی واحدها به روز رسانی می‌شوند. احتمال شمول جدید که وابسته به $I_1,\dots,I_{j-1}$ است، را به‌صورت $p_i^{(j-1)}$ نمایش می‌دهیم. بنابراین واحد j ام با احتمال به روز شده‌ی $p_i^{(j-1)}$ وارد نمونه خواهد شد.
برای بدست آوردن فرمول کل احتمال به روز شده‌ی شمول فرض می‌کنیم،که برای $i\geq 1$ ، $p_i^{(0)}=p_i,$ باشد. بنابراین، احتمال شمول به صورت زیر به روز رسانی می‌شود:

$$p_i^{(j)}=p_i^{(j-1)}-(I_j-p_i^{(j-1)})w_{i-j}^{(j)},\ \ i=j+1,\dots,j=1,\dots,N-1 \tag{۱.۲}$$

از آنجا که در هر مرحله، مقداری از احتمال کم و یا به آن اضافه می‌شود، برای تضمین $0\leq p_i^{(j)}\leq 1$ ، لازم است وزن‌ها در رابطه‌ی زیر صدق کنند:

$$-\min\left(\frac{1-p_i^{(j-1)}}{1-p_j^{(j-1)}},\frac{p_i^{(j-1)}}{p_j^{(j-1)}}\right)\leq w_{i-j}^{(j)}\leq \min\left(\frac{1-p_i^{(j-1)}}{1-p_j^{(j-1)}},\frac{1-p_i^{(j-1)}}{p_j^{(j-1)}}\right) \tag{۲.۲}$$

که به راحتی از قرار دادن فرمول (۱.۲) در بازه $[0,1]$ بدست می‌آید.
حال اگر وزن‌ها را طوری انتخاب شوند،که مجموع آن‌ها در هر گام به روز رسانی برابر با یک باشد، اندازه‌ی ثابت n برای حجم نمونه بدست می‌آید، اگر $n=\sum_{i=1}^{N}p_i=E(n(s))$ باشد، می‌توان آن را به صورت یک قضیه اثبات کرد.

قضیه ۱.۲. اگر مجموع وزن‌ها برابر $1=\sum_{j=i+1}^{N}w_{i-j}^{(i)}$ و همچنین $n=\sum_{i=1}^{N}p_i$ باشد، آن گاه اندازه‌ی نمونه مقدار ثابت و صحیح n خواهد بود.

اثبات. اثبات در ضمیمه آورده شده است. □

اکنون راهبردهای وزنی را که قرار است در بخش شبیه‌سازی باهم مقایسه شوند، معرفی می‌کنیم.

۱.۲ راهبرد وزن‌های بیشینه با مجموع ۱

راهبرد اول استفاده کردن از وزن‌های بیشینه‌ای است، [۵] که جمع آن‌ها برابر یک باشد. به طوری که نخستین وزن یعنی $w_1^{(j)}$ را تا جایی که امکان دارد بزرگ، و بقیه وزن‌ها را تا حد امکان بیشینه در کران‌های مجاز انتخاب می‌کنیم ، با این محدودیت که وزن‌ها غیر منفی با مجموع برابر یک باشند. بنابراین از مقدار وزن‌های $w_k^{(j)}$ کاسته می‌شود، وقتی k افزایش پیدا می‌کند، زیرا مجموع وزن‌ها باید برابر یک باشد. این کاهش مقدار ممکن است، باعث صفر شدن مقدار وزن‌ها برای واحدهای پایانی شود.
مشکلی که راهبرد وزن‌های بیشینه دارد، این است که اگر وزن‌ها بیشترین مقدار ممکن خود را اختیار کنند ولی مجموع آن‌ها کمتر از یک باشد، دیگر در شرایط قضیه ۱.۲ صدق نمی‌کند و حجم نمونه ثابت نخواهد شد. چون امکان این که مقدار وزن‌های بیشینه را افزایش دهیم تا مجموع آن‌ها یک شود، نیست. بنابراین برای حل این مشکل در این مقاله راهبرد جدید میانگین وزن‌های بیشینه ارائه شد، که به صورت زیر تعریف می‌شود.

۲.۲ راهبرد میانگین وزن‌های بیشینه

در این راهبرد نخستین وزن یعنی $w_1^{(j)}$ را تا حد امکان بزرگ و همچنین بقیه وزن‌ها را تا حد امکان بیشینه در کران‌های مجاز انتخاب می‌کنیم. سپس هر وزن را تقسیم بر مجموع کل وزن‌ها می‌کنیم. بنابراین وزن‌های جدید به صورت زیر خواهد بود:

$$w_{i-j}^{*(j)}=\frac{w_{i-j}^{(j)}}{\sum_{i=j+1}^{N}w_{i-j}^{(j)}} \tag{۳.۲}$$

که در آن $1 = \sum_{i=j+1}^{N} w_{i-j}^{*(j)}$. اگر مجموع وزن‌ها بیشینه $w_{i-j}^{(j)}$ برابر یک و یا، این امکان باشد، که وزن‌ها طوری انتخاب شوند که مجموع آن‌ها برابر یک شود، آن‌گاه $w_{i-j}^{*(j)}$ همان وزن‌های بیشینه $w_{i-j}^{(j)}$ خواهند بود. این نشان می‌دهد، راهبرد وزن‌های بیشینه حالت خاصی از راهبرد میانگین وزن‌های بیشینه است.

۳.۲ راهبرد وزن‌های برابر با مجموع ۱

راهبرد سوم ، راهبرد وزن‌های برابر با مجموع یک است، [۵] که در شرایط زیر صدق کند:

۱. برای $k \leq m$ قرار می‌دهیم: $w_k^{(j)} = \frac{1}{m}$.

۲. برای $k > m$ قرار می‌دهیم $w_k^{(j)} = 0$. برای مثال فرض کنید، پژوهش‌گر فقط دو واحد اول در هرگام از به روز رسانی کردن احتمال شمول را بخواهد وزن دهد، بنابراین برای $k \leq 2$ خواهیم داشت: $w_k^{(j)} = \frac{1}{2}$ ، که ممکن است در همان ابتدای به‌روز کردن احتمال شمول، وزن‌ها از کران‌ها خارج شوند، در نتیجه راهبرد وزن‌دهی برابر نیاز به اصلاح و کامل‌تر شدن دارد. بنابراین راهبرد جدید تعمیم وزن‌های برابر، در این مقاله به صورت زیر معرفی شده است.

۴.۲ راهبرد تعمیم وزن‌های برابر

در این راهبرد، برابری وزن‌ها فقط در هرگام به‌روز رسانی الزامی است. فرض کنید M_j برداری باشد، که درایه های آن وزن‌های $w_k^{(j)}$ در گام j برای واحدهای $k = j, \ldots, N$ باشد، به طوری‌که درایه‌های غیر صفر بردار M_j باید برابر باشند و در کران‌های نامساوی (۲.۲) قرار گیرند. بنابراین وزن‌ها را برای گام‌های متفاوت به‌روز رسانی احتمال‌های شمول،می‌توان متفاوت در نظر گرفت.
اگر درایه‌های غیر صفر بردار M_j برای هر گام برابر و دارای مجموع یک باشند، راهبرد تعمیم وزن‌های برابر همان راهبرد وزن‌های برابر با مجموع یک خواهد بود.

۳ نمونه‌گیری فضایی همبسته پواسون

روش نمونه‌گیری فضایی همبسته پواسون $(SCPS)$ تعدیلی از نمونه‌گیری همبسته پواسون (CPS) است، با این تفاوت که در روش $SCPS$ نیاز به محاسبه نوعی فاصله بین واحدها است، و یا این که باید فاصله‌ی بین واحدهای نزدیک به هم معلوم باشد. از این رو تابع $L(i, j)$ را تعریف می‌کنیم، که برای تعیین فاصله‌ی بین واحدهای i ام و j ام به کار می‌رود. این فاصله را می‌توان به صورت فاصله اقلیدسی یا هر اندازه فاصله‌ی متریک محاسبه کرد. در بعضی از موارد رتبه‌ی واحدها، برای تعیین فاصله‌ی این که کدام واحد نزدیک‌تر است،کافی است. بعد از تعیین تابع فاصله، به وسیله روش موزاییک‌کاری یا چند ضلعی ورونوی جامعه به ناحیه‌های یا بخش‌های مختلف تقسیم می‌شود، به صورتی که اگر واحدهای نمونه‌گیری شده را با مجموعه‌ی $S = \{s_1, \ldots, s_n\}$ نشان دهیم، چند ضلعی ورونوی برای واحد s_i شامل جمعیت تمام واحدهای نزدیک‌تر به آن نسبت به هر واحد s_j است. اگر یک واحد جمعیت فاصله مساوی با دو یا بیشتر از دو واحد از نمونه داشته باشد، آن گاه شامل بیش از یک چند ضلعی خواهد بود و احتمال شمول آن واحد به طور مساوی بین چند ضلعی‌های که آن واحد شامل آن‌ها می‌شود ، تقسیم می‌شود.
حال فرض کنید π_i ، مجموع احتمال‌های شمول همه‌ی واحدها چند ضلعی i ام باشد. برای هر انتخاب تصادفی واحد s_i مقدار مورد انتظار $E(\pi_i) = 1$ خواهد بود، زیرا نمونه شامل n واحد است و از رابطه‌ی $\sum_{i=1}^{n} \pi_i = \sum_{i=1}^{N} p_i = n$ مقدار مورد انتظار چند ضلعی به سادگی قابل محاسبه می‌شود. بنابراین برای $\Pi = (\pi_1, \ldots, \pi_n)$ واریانس را به صورت زیر تعریف می‌کنیم:

$$var(\Pi) = \frac{1}{n}\sum_{i=1}^{n}(\pi_i - 1)^2 \qquad (۱.۳)$$

اگر $var(\Pi)$ برابر با صفر و یا به عبارتی دیگر $\pi_i = 1$ برای $i = 1, \ldots, n$ باشد، فضای نمونه‌گیری متعادل است. بنابراین نزدیکی $var(\Pi)$ به صفر معیاری متعادل بودن فضا است.
در این قسمت، سه روش برای انتخاب وزن‌ها را مورد بررسی قرار می‌دهیم.

۱.۳ راهبرد وزن‌های بیشینه فضایی

در این روش برای نزدیک‌ترین واحد به واحد j ام، بیشترین وزن در نظر گرفته می‌شود و به دومین واحد از نظر کم ترین فاصله به واحد j ام وزنی تعلق می‌گیرد که تا حد ممکن بزرگ باشد و به همین ترتیب تا واحد N ام، با این محدودیت که مجموع وزن‌ها برابر یک است. به طور بدیهی برای فواصل برابر، وزن‌ها هم برابر خواهند بود. بنابراین اولویت نخست قرارگرفتن وزن‌ها بر واحدها نزدیک به واحد j ام است. در این راهبرد اگر مجموع احتمال‌های شمول برابر با یک عدد صحیح باشد می‌توان نمونه‌ی با اندازه ثابت داشت.
یک ویژگی از راهبرد وزن‌های بیشینه فضایی، ثابت کردن اندازه‌ی نمونه در هر بخش یا ناحیه است. به طوری که اگر هر بخش به شکل طبقه‌ی باز و بدون مرز دقیق باشد، برای هر بخش مانند C ، تعداد واحدهایی که انتخاب خواهند شد، برابر با $\sum_{i\in C} p_i$ خواهد بود. این ویژگی را در نتیجه‌ی از قضیه ۱.۲ در ادامه بیان خواهیم کرد.

نتیجه ۱.۳. فرض کنید جامعه شامل دو ناحیه C و B از هم مجزا باشد، به طوری که درون ناحیه همیشه فاصله‌ی بین واحدها کمتر از فاصله‌ی بین دو واحد از ناحیه‌های مختلف باشد. اگر احتمال‌های شمول به گونه‌ی تعیین شوند که $\sum_{i\in C} p_i = n_C$ و $\sum_{i\in B} p_i = n_B$ و همچنین n_C و n_B اعداد صحیحی باشند، آن‌گاه راهبرد وزن‌های بیشینه، نمونه‌ی با اندازه‌های ثابت به ترتیب n_C و n_B تولید خواهند کرد. بنابراین بدون در نظر گرفتن ترتیب نمونه‌گیری اندازه‌ی نمونه ثابت خواهد شد.

۲.۳ راهبرد وزن‌های مقدماتی نرمال

راهبرد دیگر انتخاب وزن‌هایی است،که از توزیع نرمال با میانگین صفر پیروی می‌کنند و با توجه به فاصله واحد i از واحد j، وزن‌ها مشخص خواهند شد و همچنین مجموع آن‌ها باید برابر یک باشد،که توزیع آن‌ها می‌تواند به صورت زیر باشد:

$$w_j^{(i)} \propto \exp(-(L(i,j)/\sigma^{۲})),\ \ i = j+۱, \ldots, N. \tag{۲.۳}$$

در این جا σ پارامتری است،که می‌توان برای کنترل پراکندگی وزن‌ها استفاده شود، یک گزینه برای σ، میانگین یا میانه‌ی فاصله‌ی بین هر واحد و نزدیک‌ترین واحد همسایه باشد. [۴] طبق فرمول بالا وزن‌های بزرگ بر روی واحدهای قرار می‌گیرد ،که نزدیک‌تر به j باشد.

۳.۳ راهبرد میانگین وزن‌های بیشینه فضایی

در این راهبرد همانند آنچه در بخش دوم مقاله گفته شد، وزن‌های تعیین شده برای هر واحد به وسیله راهبرد وزن‌های بیشینه فضایی، تقسیم بر مجموع کل وزن‌ها خواهدشد. وزن‌های جدید به صورت محاسبه می شوند:

$$w_{i-j}^{*(j)} = \frac{w_{i-j}^{(j)}}{\sum_{i=j+۱}^{N} w_{i-j}^{(j)}}$$

که در آن $\sum_{i=j+۱}^{N} w_{i-j}^{*(j)} = ۱$.

۴ شبیه سازی

در این بخش با شبیه‌سازی توسط نرم افزار R ، راهبردهای وزنی که در بخش‌های قبل معرفی شد، را ابتدا در یک جامعه کوچک و سپس در جامعه بزرگ‌تر با هم مقایسه کنیم. سپس نشان داه خواهد شد که راهبردهای وزنی جدید و راهبردهای وزنی پیشین باعث کم شدن اریبی برآورد پارامتر جامعه و همچنین کم شدن تغییرات حجم نمونه نسبت به نمونه‌گیری پواسون می‌شود. همچنین با راهبردهای وزنی جدید معرفی شده در این مقاله نتایج بهتری نسبت به راهبردهای پیشین خواهیم داشت، که برای مقایسه از معیارهای زیر استفاده می‌کنیم.

$$n_R = \frac{\sum_{s=۱}^{R} n(s)}{R}, \quad MR = MSE(R) = \frac{۱}{R}\sum_{s=۱}^{R}(Y_{HT} - Y_T)^{۲} \tag{۱.۴}$$

$$\hat{Y}_{HT}(s) = \sum_{i\in S} \frac{y_i}{p_i} \tag{۲.۴}$$

که n_R میانگین اندازه نمونه در $R = ۱۰۰۰$ بار تکرار نمونه‌گیری و MR میانگین مربع خطا برای هر نمونه s در $R = ۱۰۰۰$ بار شبیه‌سازی است. $\hat{Y}_{HT}(s)$ شکل نمونه‌ای برآوردگر هورویتز –تامپسون است.

جدول ۱: نتایج شبیه‌سازی برای مثال اول

راهبرد ها	mean (n)	std (n)	MR	mean $(\hat{Y}_R)$
وزن‌های بیشینه	۰۶/۲	۴/۰	۵۱/۱	۲/۱۴
میانگین وزن‌های بیشینه	۲	۰	۵۶/۰	۳/۱۵
وزن‌های برابر	۲/۲	۳۱/۰	۵۴/۲	۱۶/۱۲
تعمیم وزن‌های برابر	۲	۰	۰۲/۲	۹۲/۱۴
وزن‌های صفر یا پواسون	۰۵/۳	۰۹/۱	۰۱/۵	۱/۱۸

مثال **۱.۴**. یک جامعه با حجم $N = ۱۰$ و مجموع $Y_T = \sum_{i=۱}^{۱۰} y_i = ۱۵$ را در نظر بگیرید، به طوری که واحدهای جامعه به صورت $y = (۳, ۲, ۳, ۲, ۲, ۱, ۰, ۱, ۰, ۱)$ و متغیرهای کمکی متناسب با واحدها به صورت بردار $x = (۱۰, ۱۵, ۱۵, ۱۵, ۱۵, ۱۵, ۱۵, ۱۵, ۱۵, ۷۰)$ نشان داده می‌شود. احتمال شمول را با فرض این که $c = ۰/۰۱$ است، متناسب با متغیر کمکی به صورت $p_i = cx_i$ تعریف می‌کنیم. بنابراین $\sum_{i=۱}^{۱۰} cx_i = ۲ = n \sum_{i=۱}^{۱۰} p_i$. همان طور که در جدول خواهیم دید، راهبرد وزن های بیشینه، اندازه نمونه ثابت را نتیجه نخواهد داد. چون مجموع بیشترین مقدار وزن آن‌ها کمتر از یک شده است. راهبرد وزن‌های برابر با مجموع یک را به صورت زیر تعریف می‌کنیم :

۱. برای $k \leq ۵$ قرار می‌دهیم: $w_k^{(j)} = \frac{۱}{۵}$.

۲. برای $k > ۵$ قرار می‌دهیم: $w_k^{(j)} = ۰$.

همچنین برای راهبرد تعمیم وزن‌های برابر خواهیم داشت:

$$M_j = \frac{۱}{N-j} = \frac{۱}{۱۰-j}, j = ۱, \dots, ۹.$$

نتایج بدست آمده در جدول ۱.۴ قابل مشاهده است.
همان طور که در جدول مشاهده می کنید، راهبردهای وزنی نتایج بهتری نسبت به نمونه‌گیری پواسون عادی داشته‌اند و همچنین راهبردهای جدید یعنی میانگین وزن‌های بیشینه و تعمیم وزن‌های برابر از راهبردهای پیشین خود کارایی بهتری داشته‌اند.

مثال **۲.۴**. (قدم تصادفی) در این مثال بین واحدهای جامعه y_i ها یک روند یا همبستگی مثبت وجود دارد و همچنین y_i ها از فرآیند قدم زدن تصادفی از رابطه زیر بدست می‌آیند:

$$y_i = \sum_{j=۱}^{i} z_j, \quad i = ۱, ۲, \dots, N$$

حال آن که z_i ها را از توزیع نرمال استاندارد می‌باشند و از جامعه y_i ها با حجم $N = ۱۰۰$ ، برای هر راهبرد ۱۰۰۰ بار نمونه‌گیری تکرار شده است. احتمال شمول متناسب با متغیر کمکی را با فرض این که $c = \frac{۱}{۲۰۲}$ و $x_i = i$ است، به صورت $p_i = cx_i = \frac{i}{۲۰۲}$ تعریف می‌کنیم. بنابراین برای اندازه نمونه خواهیم داشت:

$$\sum_{i=۱}^{۱۰۰} cx_i = ۲۵ = n \sum_{i-۱}^{۱۰۰} p_i.$$

از آن جای که مقادیر بزرگ y_i ها که در مجموع متغیر واحدها تاثیر بیشتری دارند، در گام‌های پایانی بدست می‌آیند، بنابراین احتمال شمول در واحدهای پایانی را طوری انتخاب می‌کنیم که بزرگتر از احتمال شمول واحدهای ابتدای باشند. راهبرد وزنی بیشینه و میانگین راهبرد وزن های بیشینه و دو راهبرد وزن‌های برابر برای مقایسه در این مثال استفاده می‌شود، دو راهبرد وزن‌های برابر به صورت زیر است:

۱. راهبرد تعمیم وزن های برابر برای هر مرحله j ام به روز رسانی، وزن‌ها را واحدها برابر با $w_i^{(j)} = M_j = \frac{۱}{۱۰۰-j}$ قرار می‌دهیم که در آن $i = j+۱, \dots, ۱۰۰$ و $j = ۱, \dots, ۹۹$ است .

۲. راهبرد وزن‌های برابر با مجموع یک را به صورت زیر تعریف می‌کنیم :

جدول ۲: نتایج مثال دوم برای ۱۰۰۰ بار تکرار

راهبرد ها	mean (n)	std (n)	MR	mean $(\hat{Y}_R)$
وزن‌های بیشینه	۰۱/۲۵	۳۴/۰	۴۲/۲	۲/۵۵۱
میانگین وزن‌های بیشینه	۲۵	۰	۷۱/۱	۱/۵۵۴
وزن‌های برابر	۲/۲۴	۸۱/۰	۹۱/۲	۴/۵۵۰
تعمیم وزن‌های برابر	۲۵	۰	۰۱/۲	۹/۵۵۳
وزن‌های صفر یا پواسون	۰۵/۲۷	۷۹/۲	۵۱/۶	۲/۵۶۰

جدول ۳: نتایج مربوط به مثال سوم

Strategy	mean (n)	std (n)	MR	mean $(\hat{Y}_R)$	balance Spatial
۱SCPS	۸/۱۱	۰۱/۰	۶۴/۵	۵۷/۰	۱۳۴/۰
۲SCPS	۹/۱۰	۱/۲	۵۲/۵	۶۴/۰	۱۳۴/۰
۳SCPS	۱۲	۰	۷۸/۵	۲۶/۰	۱۳۴/۰
SRS	۱۲	۰	۴۵/۵	۷۶/۱	۳۰۹/۰

(آ) برای $k \leq ۲۵$ قرار می‌دهیم: $w_k^{(j)} = \frac{۱}{۲۵}$.

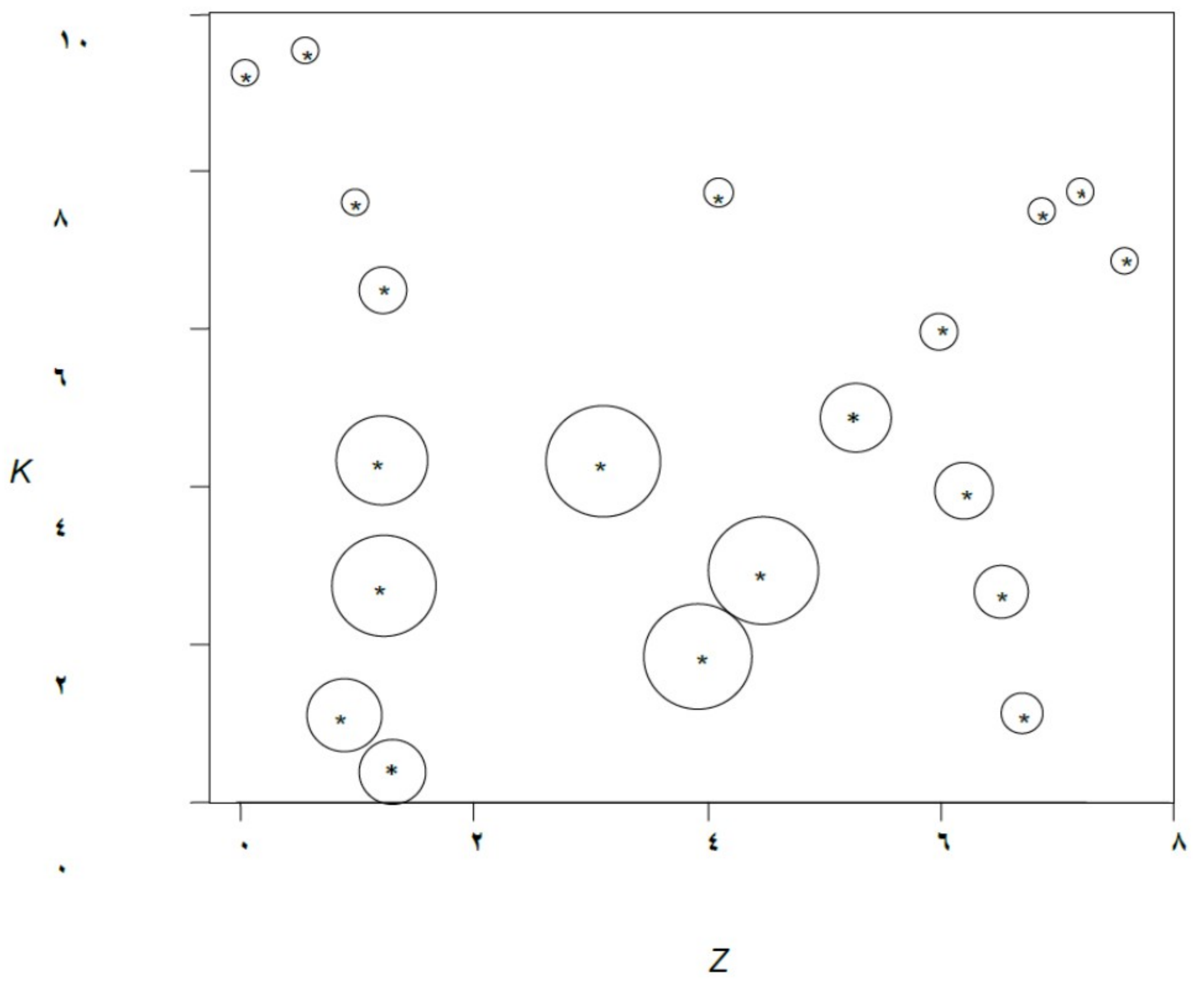
(ب) برای $k > ۲۵$ قرار می دهیم: $w_k^{(j)} = ۰$

نتایج را در جدول زیر برای راهبردهای مختلف نشان می‌دهیم: از آنجا که مجموع واحدهای جامعه برابر با ۵۵۳/۹۵ است و با توجه به نتایج بدست آمده، راهبرد میانگین وزن‌های بیشینه و تعمیم وزن‌های برابر برآورد نزدیک‌تری به مجموع واحدها و تغییرات کمتری در اندازه نمونه و میانگین مربع خطا دارند. اما در کل راهبردهای وزنی نتایج بهتری نسبت به نمونه‌گیری پواسون عادی داشته اند.

مثال **۳.۴**. جامعه ی به حجم $N = ۲۰$ را فرض کنید، که اعضای جامعه در شکل ۱ به صورت مرکز دایره مشخص شده اند. در این شکل مرکز هر دایره که با علامت ضرب مشخص شده موقعیت واحدها را نشان می‌دهد و ناحیه درون دایره‌ها فضایی از اعضای جامعه است، که با متغیر مورد مطالعه y_i واحد i ام رابطه یا روند قوی دارد. بنابراین قبل از انجام نمونه‌گیری اطلاعات ما فقط در مورد موقعیت واحدها است. احتمال شمول را برای واحدها برابر در نظر می‌گیریم. در این مثال طبقه‌های فضایی به روشنی قابل تشخیص نیست و هدف برآورد مجموع متغیرهای y_i است، به طوری که هر واحد i مقدار صفت y_i را در بر می‌گیرد، که این صفت را سطح دایره‌ی متناظر با واحد i ام در نظر می‌گیریم. در این مثال برآوردگر هورویتز–تامپسون استفاده شده و احتمال شمول برابر با $p_i = \frac{۳}{۵}$ و متغیر کمکی $x_i = ۱$ برای هر واحد i است. مجموع سطوح برابر با ۵/۷۹ است. طرح نمونه‌گیری $SCPS۱$ با وزن‌های بیشینه و $SCPS۲$ با وزن‌های ابتدایی نرمال با پارامتر $\sigma = ۲$ و طرح نمونه‌گیری $SCPS۳$ میانگین وزن‌های بیشینه فضایی و همچنین نمونه‌گیری تصادفی ساده (SRS) با حجم نمونه ای ۱۲، چهار روشی است، که برای مقایسه استفاده می‌کنیم. داده‌های مربوط به جامعه در جدول ۵ که در ضمیمه موجود می‌باشد، نشان داده شده است و جدول ۳.۴ نتایج مربوط به روش‌های مختلفی است که استفاده کرده‌ایم . نتایج بدست آمده با هزار بار تکرار نمونه‌گیری حاصل شده است، که ستون‌های جدول ۳.۴ از طریق رابطه (۲.۳) محاسبه شده و اندازه ی فضایی متعادل با توجه به رابطه (۱.۳) محاسبه شده است. هر چه این اندازه کمتر باشد ، درجه ی بالایی را از فضای متعادل شده را نشان می‌دهد. با توجه به داده‌های جدول ۳.۴ می توان نتیجه گرفت که روش‌های نمونه‌گیری فضایی عملکرد بهتری نسبت به نمونه‌گیری تصادفی ساده دارند. در مجموع استراتژی میانگین وزن‌های بیشینه فضایی دارای MR یا MSE کمتری نسبت به دیگر روش‌ها است و برآورد نزدیک تری به پارامتر جامعه دارد.

۵ نتیجه‌گیری

نمونه‌گیری همبسته پواسون چه در حالت تک بعدی یا فضایی برای جوامعی که بین واحدهای مجاور آن‌ها همبستگی مثبت وجود دارد، تا کنون ارائه و بررسی شده است. بدین صورت که با ایجاد همبستگی منفی بین تابع شاخص شمول واحدهای مجاور و کاهش احتمال شمول واحدهای مجاور باعث کاهش همبستگی مثبت و اریبی برآورد پارامتر جامعه و تغییرات اندازه نمونه می‌شود. این همبستگی منفی به کمک راهبردهای وزنی ایجاد می‌شود، به طوری که محقق در انتخاب راهبردهای وزنی آزاد است، به شرط این که این راهبردهای وزنی اصل اساسی را که احتمال شمول واحدها باید بین صفر و یک باشد را نقض نکند. بنابراین راهبردهای وزنی جدید زیادی را می‌توان معرفی و بررسی کرد.



شکل ۱: دایره‌ها مربوط به فضایی است، که واحدها روند قوی با واحد جامعه یا مرکز دایره دارند

پیوست

اثبات قضیه ۱. طبق فرض اگر وزن‌ها را طوری انتخاب شوند ،که مجموع آن‌ها در هر گام به روز رسانی برابر با یک باشد، یعنی :

$$\sum_{j=i+1}^{N} w_{i-j}^{(i)} = 1$$

و اگر دو طرف رابطه به روز رسانی واحد j ام در گام i ام را به صورت زیر مجموع ببندیم، خواهیم داشت:

$$\sum_{j=i+1}^{N} p_j^{(i)} = \sum_{j=i+1}^{N} \left(p_j^{(i-1)} - (I_i - p_i^{(i-1)}) w_{j-i}^{(i)}\right)$$

با فرض مجموع یک برای وزن‌ها به رابطه‌ی زیر می‌رسیم :

$$\sum_{j=i+1}^{N} p_j^{(i)} = \sum_{j=i+1}^{N} p_j^{(i-1)} - I_i.$$

با فرض $p_j^{(0)} = p_j$ و قرار دادن مقادیر $1, 2, \dots, t$ بجای i در رابطه‌ی (۱.۳) به طور متوالی و به صورت بازگشتی به رابطه‌ی زیر می‌رسیم :

$$\sum_{j=t+1}^{N} p_j^{(t)} = \sum_{j=1}^{N} p_j - (I_1 + \dots + T_t).$$

اکنون با در نظر گرفتن $t = N - 1$ رابطه‌ی زیر را خواهیم داشت :

$$\pi_N^{(N-1)} = \sum_{j=1}^{N} \pi_j - (I_1 + \dots + I_{N-1}).$$

با توجه به این که احتمال بین صفر و یک است، بنابراین

$$0 \leq \pi_N^{(N-1)} = n - (I_1 + \dots + I_{N-1}) \leq 1.$$

پس $\sum_{j=1}^{N-1} I_j$ برابر n و یا $n-1$ خواهد بود. چون عبارت $\sum_{j=1}^{N} I_j - n$ یک عدد صحیح مثبت است، $\pi_N^{(N-1)}$ یا صفر یا یک می‌شود، یعنی خود متغیر تصادفی است که مقدارش برابر با I_N است. اگر $\pi_N^{(N-1)}$ برابر با یک باشد و مجموع مابقی توابع شمول برابر با $n-1$ باشد ، اندازه‌ی نمونه برابر با n خواهد شد و اگر $\pi_N^{(N-1)}$ برابر با صفر باشد، مجموع مابقی توابع شمول برابر با n می‌شود و اندازه‌ی نمونه دوباره برابر با n خواهد شد.

جامعه مورد بررسی مثال ۳

جدول زیر مربوط داده‌های جامعه مورد بررسی در مثال سوم است، که شماره هر واحد به ترتیب در ستون اول قرار گرفته و موقعیت مکانی در ستون های دوم و سوم و صفت مورد بررسی (متغیر هدف) در ستون آخر قرار گرفته است.

۱. Sampling Poisson Correlated

۲. Sampling Poisson Correlated Spatially

جدول ۴: داده های جامعه برای مثال سوم

Target variable(y)	k-position	z-position	Unit
۱۹/۰	۱۲/۱	۷۳/۶	۱
۱۱/۰	۸۴/۷	۳۰/۴	۲
۶۵/۰	۹۲/۲	۵۲/۴	۳
۱۷/۰	۰۴/۶	۱۰/۶	۴
۰۱/۰	۶۴/۹	۵۹/۰	۵
۶۸/۰	۳۲/۴	۱۶/۳	۶
۰۴/۰	۹۵/۶	۷۳/۷	۷
۰۴/۰	۵۸/۷	۹۶/۶	۸
۵۳/۰	۳۳/۴	۲۵/۱	۹
۲۳/۰	۵۵/۶	۳۰/۱	۱۰
۴۲/۰	۱۰/۱	۹۲/۰	۱۱
۰۲/۰	۳۴/۹	۰۸/۰	۱۲
۶۳/۰	۸۷/۱	۲۳/۴	۱۳
۲۸/۰	۶۶/۲	۵۶/۶	۱۴
۰۳/۰	۹۸/۷	۲۳/۷	۱۵
۳۹/۰	۸۸/۴	۳۱/۵	۱۶
۱۰/۰	۶۹/۷	۰۹/۱	۱۷
۳۰/۰	۹۶/۳	۳۲/۶	۱۸
۶۱/۰	۷۳/۲	۲۶/۱	۱۹
۳۶/۰	۳۷/۰	۳۴/۱	۲۰

فهرست منابع

[1] Altieri, L., Cocchi, D., Spatial Sampling for Non-compact Patterns, International Statistical Review, (2021), **89,** 532-549.

[2] Bondesson, L., Thorburn, D., A list sequential sampling method suitable for realtime sampling, Scand. J. Statist., (2008), **35**, 466–483.

[3] Grafstrom, A., Matei,A, Spatially Balanced Sampling of Continuous Populations, (2018), Scand. J. Statist., (2018), **45**, 792 –805.

[4] Grafstrom, A, Spatially correlated Poisson sampling, J. Statist. Plann. Inference, (2012), **142,** 139–147.

[5] Grafstrom, A., On a generalization of Poisson sampling, J. Statist. Plann. Inference, (2010), **140,** 982–991.

[6] Hajek, J. (1964). Asymptotic theory of rejective sampling with varying probabilities from a finite population. Annals of Mathematical Statistics, 35, 1491–1523.

[7] Horvitz, D.G. and Thompson, D.J. (1952). A generalization of sampling without replacement from a finite universe. Journal of the American Statistical Association, 47, 663–685.

[8] Meister, K. (2004). On methods for real time sampling and distributions in sampling. Ph.D. thesis, Department of Mathematical Statistics, Umea University, Umea.

[9] Ogus, J. L. and Clark, D.F. The annual survey of manufactures: A report on methodology, (1971). Technical paper no. 24. Bureau of the Census, Washington D.C.

[10] Sarndal, C.E., Swensson, B. and Wretman, J.H., Model Assisted Survey Sampling. (1992), New York: Spinger-Verlag.

New weighting strategies for spatially correlated Poisson sampling

H. Farokhinia [†], **R. Chinipardaz, G.A. Parham**

Department of Statistics, Faculty of Mathematics and Computer Science, Shahid Chamran University of Ahvaz, Ahvaz, Iran.

Communicated by: G.R. Mohtashami Borzadaran

Received: 2022/6/13 Accepted: 2022/12/6

Abstract: In Poisson sampling, each unit is selected independently of the other units with a certain inclusion probability and the sample size n(s) is a random variable with large variation. Also, if there exists a trend between units of the population, it causes bias in the estimated population parameter. So Poisson-correlated sampling (CPS) and Poisson-correlated spatial sampling (SCPS) were introduced as alternative methods to Poisson sampling that reduces changes in the sample size and bias in the estimated population parameter by weighting strategies. In this paper, new strategies for choosing weights are introduced, and it is shown by simulation that the new weighting strategies increase the efficiency of the estimated parameter compared to the earlier weighting strategies.

Keywords: Correlated Poisson sampling, Horvitz–Thompson estimator, Poisson sampling, Spatial sampling.

[†] **Corresponding author:** hadi-farohkinia@stu.scu.ac.ir