یک طرح نمونه‌گیری پواسون کارا در حالت متناسب با اندازه

پگاه افشین، بردیا پناه‌بحق و امیرحسین صنعت‌پور

دانشکده علوم ریاضی و کامپیوتر، دانشگاه خوارزمی

چکیده: در این مقاله یک روش اصلاح‌شده نمونه‌گیری پواسون با حداقل حجم ثابت نمونه ارائه شده است. این طرح ترکیبی از روش‌های نمونه‌گیری پواسون و نمونه‌گیری تصادفی ساده بدون جایگذاری است. از نمونه‌گیری تصادفی ساده برای جبران حجم نمونه از اعضای باقیمانده جامعه، بعد از نمونه‌گیری پواسون استفاده می‌شود. در مرحله اول، هر واحد بر اساس طرح پواسون به‌طور مستقل و با توجه به احتمال شمولی که از قبل تعیین‌شده، انتخاب و درصورت نرسیدن به یک حجم نمونه از پیش تعیین‌شده، در مرحله دوم از نمونه‌گیری تصادفی ساده برای جبران کمبود نمونه استفاده می‌شود. از مزایای این طرح می‌توان به سادگی در اجرا، کنترل حجم نمونه، توانایی اجرای روش احتمال تقریباً متناسب با اندازه، و کارایی بیشتر نسبت به سایر طرح‌های اصلاح‌شده پواسون در حضور متغیرهای کمکی با همبستگی متوسط اشاره کرد.

واژه‌های کلیدی: نمونه‌گیری پواسون اصلاح‌شده، کارایی، احتمال متناسب با اندازه.

°آدرس الکترونیک مسئول مقاله: پگاه افشین، pegah.afshin@gmail.com

°کد موضوع‌بندی ریاضی (۲۰۱۰): 62D05، 62N02، 62L05.

۱ مقدمه

نمونه‌گیری فهرست‌دنباله‌ای[1] به مجموعه‌ای از روش‌های نمونه‌گیری گفته می‌شود که در آن‌ها با مرور کل یا دنباله‌ایِ فهرست یا چارچوب اعضای جامعه، نمونه مد نظر انتخاب می‌شود. از جمله این روش‌ها، نمونه‌گیری پواسون[2] است که در آن ابتدا به هر یک از اعضای جامعه شانسی برای حضور در نمونه (احتمال شمول) تخصیص داده می‌شود؛ سپس برای هر عضو، با استفاده از توزیع یکنواخت در بازه (۰, ۱)، یک متغیر تصادفی تولید شده؛ در نهایت با مرور کل فهرست اعضا، هر عضو مستقل از دیگر اعضا شانس حضور در نمونه نهایی را بر اساس یک آزمایش برنولی خواهد داشت که در آن پیروزی به‌صورت کوچکتر بودن متغیر تصادفی یکنواخت از احتمال شمول مربوطه تعریف می‌شود.

از معایب این طرح نمونه‌گیری تغییرپذیری زیاد حجم نمونه و امکان مواجهه با حجم نمونه صفر است. یکی از اولین ایده‌ها برای دوری از حجم نمونه صفر یا رسیدن به یک حداقل حجم نمونه، تعیین امید حجم نمونه CPS به گونه‌ای است که احتمال رسیدن به حجم نمونه صفر یا کوچک، بسیار کم باشد. اما به‌دلیل وجود احتمال مثبت برای این حالت‌ها و همچنین عدم کنترل رشد حجم نمونه که منجر به تغییرپذیری زیاد حجم نمونه می‌شود، این ایده از مطلوبیت کافی برای استفاده در مسائل کاربردی برخوردار نیست . از مزیت‌های CPS می‌توان به استقلال انتخاب نمونه‌ها از یکدیگر، که امکان انجام یک نمونه‌گیری متناسب با اندازه بدون جایگذاری را فراهم می‌آورد اشاره کرد (سارندال و همکاران، ۲۰۰۳). طرح‌های متناسب با اندازه به طرح‌هایی اطلاق می‌شوند که در آن‌ها احتمال شمول اعضا متناسب با متغیر مد نظر است. این طرح‌ها درصورت وجود متغیرهای کمکی مناسب (با هزینه کم و همبستگی بالا) طرح‌هایی بسیار کارا هستند اما اجرای آن‌ها در نمونه‌گیری‌هایی بدون جایگذاری که در آن‌ها شانس انتخاب اعضا به یکدیگر وابسته است، بسیار پیچیده خواهد شد. نمونه‌گیری پواسون برای اولین بار توسط هایک (۱۹۵۷) به‌عنوان یک طرح نمونه‌گیری با احتمالات شمول نابرابر برای غلبه بر این پیچیدگی‌ها ارائه شد. به‌دلیل استقلال که موجب محاسبه آسان واریانس‌ها و نیز سادگی در اجرا می‌شود، این روش همواره مورد توجه آماردانان بوده است. اما نمونه‌گیری پواسون دارای ضعف بزرگ تغییرپذیری زیاد حجم نمونه است. حجم نمونه در نمونه‌گیری پواسون متغیری تصادفی و دامنه تغییرات آن از صفر تا تعداد کل جامعه است. به این دلیل برخی از پژوهشگران همواره سعی در ارائه روش اصلاح‌شده‌ای داشته‌اند که موجب کاهش

[1] List sequential

[2] Conventional Poisson sampling (CPS)

تغییرپذیری زیاد حجم نمونه شود.

اولین روش‌های اصلاح‌شده عمدتاً مبتنی بر روش‌های رد نمونه به‌دست آمده و تکرار نمونه‌گیری پواسون تا رسیدن به نمونه‌ای با حجم غیر صفر یا حجم دلخواه بود (آگوس و کلارک، ۱۹۷۱؛ برووِر و همکاران، ۱۹۸۰ و قوش و وت، ۱۹۹۸). این روش‌ها با توجه به تغییرات طرح نمونه‌گیری، به روش‌هایی آسان از نظر اجرا ولی پیچیده از نظر محاسبات احتمالات شمول تکی و توام (احتمال انتخاب دو عضو خاص به‌طور همزمان در نمونه) تبدیل شدند. روش‌های اصلاحی مطرح شده بعدی تحت عنوان نمونه‌گیری پواسون همبسته، بر اساس طرح احتمالات نابرابر با استفاده از روش خرد کردن احتمالات شمول (دویل و تیله، ۱۹۹۸ و تیله، ۲۰۰۶) دنبال شد. در روش خرد شده، بردار احتمالات شمول به چند بردار احتمالات تقسیم و با انتخاب تصادفی یکی از این بردارها، مساله انتخاب اعضا بر روی بردار احتمالات ساده‌تری اجرا و دنبال می‌شود. در هر مرحله، پس از ساده کردن بردار احتمالات شمول، و انتخاب عضوی بر اساس آن، احتمال شمول عضو مد نظر از بردار حذف شده و مساله خرد کردن بر روی بردار کوچک‌تر و به شکل ساده‌تری دنبال می‌شود. بر اساس این ایده، در نمونه‌گیری پواسون همبسته، با مرور فهرست اعضا، پس از عبور از هر عضو و مشخص شدن وضعیت انتخاب آن، بقیه احتمالات شمول به‌روز می‌شوند که با این به‌روزرسانی و ایجاد همبستگی انتخاب اعضا، کنترل بر روی حجم نمونه بیشتر می‌شود (بوندسن و توربارن، ۲۰۰۸ و گرفاستروم، ۲۰۱۰).

گرفاستروم (۲۰۱۰)، آنتروپی[۳] (اندازه تصادفی بودن نمونه) نمونه‌گیری پواسون همبسته را با طرح‌های مختلف مورد مقایسه و بحث قرار داد. گرفاستروم (۲۰۱۲)، طرح پواسون با همبستگی فضایی را برای هرچه بیشتر پخش کردن نمونه در جوامع یک، دو یا چند بعدی و سپس دو طرح پواسون با مختصات همبسته منفی و مثبت را به ترتیب برای پراکندگی بیشتر نمونه و کاهش هزینه جمع‌آوری نمونه معرفی نمود (گرفاستروم، ۲۰۱۵). همچنین در یکی از آخرین پژوهش‌ها، تیله و ویلهلم (۲۰۱۷) حالت مقید نمونه‌گیری پواسون با کنترل حجم نمونه را تحت سه اصل تصادفی‌سازی (دادن شانس به تمامی اعضای جامعه)، بیش‌نمایندگی (دادن شانس بیشتر برای انتخاب اعضای مهم‌تر) و محدودیت (پرهیز از انتخاب نمونه‌های بد) مورد بررسی بیشتر قرار دادند . از میان روش‌های نمونه‌گیری پواسون همبسته، روش وزن‌های بیشینه[۴] (گرفاستروم، ۲۰۱۰) دارای کارایی بالاتر و تغییرات نمونه بسیار کمی است . در این مقاله نیز روشی، با ترکیب نمونه‌گیری پواسون و تصادفی ساده، برای کنترل بیشتر و دستیابی به حداقلی از حجم نمونه مطرح خواهد شد و کارایی این روش جدید

[3] Entropy

[4] Maximal weights

در طرح‌های نمونه‌گیری متناسب با اندازه، برای همبستگی‌های متفاوت با روش وزن‌های بیشینه در پواسون همبسته مقایسه و مورد ارزیابی قرار خواهد گرفت . در بخش ۲ روش جدیدی را مطرح خواهیم کرد و بخش ۳ شامل شبیه‌سازی‌هایی بر اساس دو توزیع نرمال و پواسون و یک مجموعه داده واقعی از مطالعه بر روی گیاهان دارویی است. مقاله در بخض ۴ با بحث و نتیجه‌گیری به پایان می‌رسد.

۲ روش نمونه‌گیری پواسون اصلاح‌شده با حداقل حجم ثابت نمونه

جامعه‌ای به حجم N به‌صورت $U = \{1, \dots, N\}$ را که در آن N معلوم است، در نظر بگیرید. همچنین فرض کنید، برای $k = 1, \dots, N$، y_k مقدار متغیر مورد نظر مربوط به واحد k-ام جامعه، x_k مقدار متغیر کمکی متناظر با آن و هدف برآورد مقدار کل جامعه به‌صورت $t = \sum_{k \in U} y_k$ باشد . در روش نمونه‌گیری پواسون، ابتدا به هر یک از اعضای جامعه یک احتمال شمول (π_k برای عضو k-ام جامعه) بر اساس متغیر کمکی تخصیص داده می‌شود. سپس به تعداد اعضای جامعه، متغیر تصادفی مستقل با توزیع یکنواخت (ε_k برای عضو k-ام جامعه) در بازه $(0, 1)$ انتخاب می‌شوند. در نهایت عضو k-ام جامعه در نمونه نهایی قرار خواهد گرفت اگر $\varepsilon_k < \pi_k$. بدین‌صورت به یک نمونه تصادفی با حجم متغیر رسیده که در آن احتمالات انتخاب تمامی اعضا از یکدیگر مستقل هستند. اما روش نمونه‌گیری پواسون اصلاح‌شده[5] از دو مرحله تشکیل شده است. ابتدا از جامعه به روش نمونه‌گیری پواسون نمونه‌ای گرفته می‌شود. فرض کنید پس از این مرحله نمونه‌ای به حجم n_{po} به‌دست آمده باشد. همچنین فرض کنید هدف رسیدن به حداقل $n_0 > 0$ واحد نمونه است. بنابراین پس از اجرای این مرحله، مقدار n_0 با n_{po} مقایسه می‌شود.

اگر $n_{po} \geq n_0$، یعنی مقدار حداقل حجم نمونه مورد نظر به‌دست آمده است و نمونه‌گیری متوقف می‌شود. اما اگر $n_{po} < n_0$، نمونه‌ای به اندازه $n_0 - n_{po}$ از باقیمانده اعضا جامعه که در مرحله اول انتخاب نشده‌اند، به روش تصادفی ساده بدون جایگذاری انتخاب می‌شود. با این کار اطمینان حاصل می‌شود که پس از اجرای نمونه‌گیری هرگز نمونه‌ای به حجم صفر به‌دست نمی‌آید (که در نمونه‌گیری پواسون امکان رخداد دارد) و در حقیقت با ادغام این دو روش به حداقل حجم نمونه مورد نظر دست یافته می‌شود.

در روش پواسون، احتمالات شمول، قبل از انجام نمونه‌گیری تعیین می‌شوند (π_k برای عضو k-ام). روش‌های مختلفی را می‌توان برای تعیین این احتمالات در نظر گرفت. در اینجا از روش احتمال متناسب با

[5] Modified Poisson sampling (MPS)

اندازه با استفاده از متغیرهای کمکی استفاده می‌شود.

توجه به این نکته ضروری است که استفاده از روش متناسب با اندازه، در صورت وجود متغیر کمکی با همبستگی مناسب، موجب افزایش قابل توجه کارایی می‌شود (سارندال و همکاران، ۲۰۰۳). برور و هنیف (۱۹۸۳) ۵۰ مورد از این روش‌ها را فهرست کرده‌اند. چن و همکاران (۱۹۹۴) روشی ساده برای دستیابی به نمونه متناسب با اندازه بدون جایگذاری معرفی کردند و در یکی از آخرین پژوهش‌ها، اندرسن و همکاران (۲۰۱۵) روشی برای حل مشکل صفر بودن متغیر کمکی ارائه کردند. تمامی روش‌های مطرح شده برای نمونه‌گیری متناسب با اندازه، به‌طور تقریبی منجر به تولید نمونه دلخواه خواهند شد و از این نظر تفاوت چندانی با هم ندارند و تفاوت آن‌ها عمدتاً در سادگی یا پیچیدگی روش و همچنین نزدیکی احتمالات شمول واقعی به احتمالات مد نظر است. اما به هر حال اجرای طرح متناسب با اندازه بدون جایگذاری، برای حجم نمونه بزرگتر از ۲، پیچیده و دشوار است و یکی از مزیت‌های نمونه‌گیری پواسون، امکان اجرای طرح متناسب با اندازه به‌صورت بسیار ساده است، اما با ضعف تولید حجم نمونه تصادفی و گاهی صفر. روش پواسون اصلاح‌شده ادغامی از روش پواسون و تصادفی ساده است و در قسمت پواسون آن می‌توان روش کارای متناسب با اندازه را بکار برد، با این اطمینان که یک حداقل حجم ثابت نمونه به‌دست خواهد آمد. در ادامه برآوردگری نااریب برای MPS بر اساس برآوردگر هارویتز-تامسون (هارویتز و تامسون، ۱۹۵۲ و ناراین، ۱۹۵۱) ارائه می‌شود. این برآوردگر که در طرح‌های نمونه‌گیری با احتمال نابرابر برآوردگری نااریب است به‌صورت

$$\hat{t}_{\pi} = \sum_{k \in s} \frac{y_k}{\pi_k}, \qquad (۱)$$

تعریف می‌شود، که در آن s نمونه نهایی طرح را نشان می‌دهد.

قضیه ۱. در MPS برآوردگر هارویتز-تامسون برای پارامتر مجموع کل جامعه به‌صورت

$$\hat{t}_{\pi.MPS} = \sum_{k \in s} \frac{y_k}{\pi_k^*},$$

است، که در آن

$$\pi_k^* = P(k \in s) = \pi_k + (1 - \pi_k) \sum_{b=0}^{n_0 - 1} \frac{n_0 - b}{N - b} \sum_{i=1}^{\binom{N-1}{b}} \prod_{h=1}^{b} \pi_{h_i} \prod_{h=b+1}^{N-1} 1 - \pi_{h_i},$$

و π_k احتمال شمول نمونه‌گیری پواسون معمولی برای k-امین عضو جامعه، h_i معرف h-امین عضو از i-امین

ترکیب ممکن انتخاب b واحد از داخل جامعه و s مجموعه نمونه نهایی هستند. واریانس این برآوردگر به‌صورت

$$Var(\hat{t}_\pi) = \sum_{k\in U}\sum_{l\in U}(\pi^*_{kl} - \pi^*_k\pi^*_l)\left(\frac{y_k}{\pi^*_k}\right)\left(\frac{y_l}{\pi^*_l}\right), \qquad (2)$$

است که در آن احتمال شمول توأم به‌صورت

$$\begin{aligned}
\pi^*_{kl} &= P(k\in s, l\in s) = \pi_k\pi_l \\
&+ (1-\pi_k)(1-\pi_l)\sum_{b=0}^{n_0-2}\frac{n_0 - b(n_0-b-1)}{N-b(N-b-1)}\sum_{i=1}^{\binom{N-2}{b}}\prod_{h=1}^{b}\pi_{h_i}\prod_{h=b+1}^{N-2}1-\pi_{h_i} \\
&+ \pi_k(1-\pi_l)\sum_{b=0}^{n_0-2}\frac{n_0-b-1}{N-b-1}\sum_{i=1}^{\binom{N-2}{b}}\prod_{h=1}^{b}\pi_{h_i}\prod_{h=b+1}^{N-2}1-\pi_{h_i} \\
&+ \pi_l(1-\pi_k)\sum_{b=0}^{n_0-2}\frac{n_0-b-1}{N-b-1}\sum_{i=1}^{\binom{N-2}{b}}\prod_{h=1}^{b}\pi_{h_i}\prod_{h=b+1}^{N-2}1-\pi_{h_i}.
\end{aligned}$$

به‌دست می‌آید و برآوردگر نااریب واریانس نیز به‌صورت

$$\hat{V}ar(\hat{t}_\pi) = \sum_{k\in s}\sum_{l\in s}\frac{(\pi^*_{kl} - \pi^*_k\pi^*_l)}{\pi^*_{kl}}\left(\frac{y_k}{\pi^*_k}\right)\left(\frac{y_l}{\pi^*_l}\right). \qquad (3)$$

است. برای مشاهده برهان قضیه، بخش پیوست را ببینید.

در MPS اینکه چه تعداد از حداقل حجم نمونه مورد نظر در مرحله اول و به روش پواسون کلاسیک با احتمال متناسب با اندازه انتخاب و چه تعداد از روش تصادفی ساده تامین شود، اهمیت دارد. هر چقدر حجم نمونه انتخابی در مرحله اول نسبت بیشتری از حداقل حجم نمونه مورد نظر را به خود اختصاص دهد، روش پواسون اصلاح‌شده به احتمال متناسب با اندازه نزدیک‌تر است. با توجه به ساختار MPS می‌توان نتیجه‌گیری کرد که در این روش

- احتمالات شمول تقریباً متناسب با اندازه هستند.
- اگر همبستگی بین متغیر مورد مطالعه و متغیر کمکی قوی نباشد، احتمال متناسب با اندازه کارا نخواهد بود. اما ازآنجاکه MPS ترکیبی از احتمالات متناسب با اندازه و تصادفی ساده بدون جایگذاری است، می‌تواند با استفاده از مزیت طرح تصادفی ساده، ناکارامدی مربوطه را جبران کند.

برآوردگری کاراتر از برآوردگر هارویتز-تامسون

برآوردگر هارویتز-تامسون (۱) در حالت نمونه‌گیری متناسب با اندازه با حجم ثابت نمونه بسیار کارا است. اگر احتمالات شمول دقیقاً با متغیر مد نظر متناسب باشند $\hat{t}_\pi = t$ برقرار است، که به معنای برآورد بدون خطای پارامتر مد نظر با استفاده از نمونه تصادفی خواهد بود. اما این برآوردگر در حالت حجم متغیر نمونه، ازآنجاکه در برآورد حجم واقعی جامعه می‌تواند دچار خطا شود، به خوبی عمل نمی‌کند. به‌طور مثال در طرح نمونه‌گیری پواسون، فرض کنید π_k دقیقاً متناسب با اندازه متغیر مد نظر باشند یا به‌عبارت دیگر $\pi_k = \frac{ny_k}{\sum\limits_{k\in U} y_k}$ و $n = E(n_s)$، که در آن n_s حجم تصادفی نمونه نهایی است. دراین‌صورت، رابطه

$$\hat{t}_\pi = \sum_{k\in s} \frac{y_k}{\pi_k} = \sum_{k\in s} \frac{y_k}{\frac{ny_k}{\sum\limits_{k\in U} y_k}} = \sum_{k\in s} \frac{\sum\limits_{k\in U} y_k}{n} = \frac{n_s}{n} \sum_{k\in U} y_k = \frac{n_s}{n} t, \qquad (۴)$$

برقرار است. همان‌طور که ملاحظه می‌شود برخلاف طرح‌هایی با حجم ثابت نمونه، حتی با تعیین احتمالات دقیقا متناسب با اندازه نیز قادر به برآورد بدون خطای پارامتر مد نظر نیست. برای رفع این مشکل با توجه به (۴)، برآوردگر

$$\hat{t}_{alt.po} = \begin{cases} \frac{n}{n_s}\hat{t}_\pi & n_s > ۰ \\ ۰ & n_s = ۰ \end{cases}$$

معرفی می‌شود. همچنین بر اساس $\hat{t}_{alt.po}$ برآوردگر طرح MPS نیز به‌صورت

$$\hat{t}_{alt.MPS} = \frac{E(n_{MPS})}{n_{MPS}} \hat{t}_{\pi.MPS}, \qquad (۵)$$

اصلاح می‌شود، که در آن n_{MPS} حجم تصادفی نمونه نهایی و $E(n_{MPS}) = \sum\limits_{k\in U} \pi_k^*$ امید این حجم تصادفی نمونه در MPS است. قابل توجه است که این امید با توجه به معلوم بودن تمامی احتمالات شمول بر اساس متغیر کمکی مقداری معلوم است. با تکنیک خطی‌سازی تیلر (سارندال و همکاران، ۲۰۰۳) می‌توان نشان داد که برآوردگر (۵) به‌طور تقریبی ناا‌ریب است. در $\hat{t}_{alt.MPS}$ برخلاف $\hat{t}_{alt.po}$، حجم نمونه ظاهر شده در مخرج برآوردگر صفر نخواهد شد، زیرا همواره رابطه $n_{MPS} \geq n_۰$ برقرار است.

در حالت اجرای طرح متناسب با اندازه با اطلاعات کمکی کامل (با همبستگی یک)، اگر π_k^* تقریباً

متناسب با اندازه باشد، آنگاه

$$\begin{aligned}\hat{t}_{alt.MPS} &= \frac{E(n_{MPS})}{n_{MPS}}\hat{t}_{\pi.MPS}\\ &\simeq \frac{E(n_{MPS})}{n_{MPS}}\sum_{k\in s}\frac{y_k}{\frac{E(n_{MPS}).y_k}{t}}\\ &= \frac{E(n_{MPS})}{n_{MPS}}.\frac{n_{MPS}}{E(n_{MPS})}t = t.\end{aligned}$$

بنابراین در شبیه‌سازی‌ها از $\hat{t}_{alt.MPS}$ به‌جای $\hat{t}_{\pi.MPS}$ استفاده خواهد شد.

۳ ارزیابی روش از طریق شبیه‌سازی

برای ارزیابی عملکرد طرح MPS، این طرح را با استفاده از نرم‌افزار R (نسخه 3.5.1) و بر اساس شبیه‌سازی‌های مونت کارلو، با دو طرح زیر مورد مقایسه قرار خواهد گرفت.

- طرح نمونه‌گیری پواسون همبسته متناسب با اندازه: طرح گرفاستروم (۲۰۱۰) تحت عنوان پواسون همبسته با وزن‌های بیشینه (MW).

- طرح نمونه‌گیری متناسب با اندازه: طرح چن و همکاران (۱۹۹۴) تحت عنوان طرح متناسب با اندازه با آنتروپی بیشینه (ME).

طرح MW در طی یک نمونه‌گیری پواسون، اعضای نمونه را متناسب با اندازه یک متغیر کمکی انتخاب خواهد کرد و در طرح ME نمونه‌ای به حجم n، متناسب با اندازه یک متغیر کمکی، براساس یک فرآیند انتخاب تک تک اعضا[۶] (چن و همکاران ۱۹۹۴، صفحه ۴۶۱) انتخاب می‌شوند. بدین ترتیب، طرح MPS با هر دو حالت نمونه‌گیری پواسون متناسب با اندازه و نمونه‌گیری متناسب با اندازه ساده مقایسه خواهد شد.

برای متغیر اصلی، جامعه‌ای به حجم ۲۰۰ به‌صورت تصادفی از دو توزیع نرمال و پواسون تولید کرده و برای تولید متغیر کمکی از معادله

$$x_k = \beta y_k + \varepsilon_k, \qquad (۶)$$

[6] Draw-by-draw selection procedure

استفاده می‌شود، که در آن β یک ضریب ثابت و ε_k به‌عنوان خطا برای تولید همبستگی کوچکتر از ۱، دارای توزیع یکنواخت در بازه‌ای متقارن حول صفر است. در نهایت احتمالات شمول متناسب با اندازه به‌صورت

$$\pi_k = \frac{nx_k}{\sum\limits_{k=1}^{N} x_k},$$

مورد استفاده قرار می‌گیرند. برای ارزیابی MPS سه محک تحت عناوین کارایی (Eff) برای کارایی MPS نسبت به SRS[7]، کارایی تعدیل شده برای کارایی MPS نسبت به MW (CeffMW) و کارایی تعدیل شده برای کارایی MPS نسبت به ME (CeffME) به‌ترتیب برای مقایسه با MW و ME به‌صورت

$$\begin{aligned}
\text{Eff} &= \frac{Var_{SRS}}{MSE_{MPS}}, \\
\text{CeffMW} &= \frac{MSE_{MW}}{MSE_{MPS}} \cdot \frac{E(n_{MW})}{E(n_{MPS})}, \\
\text{CeffME} &= \frac{MSE_{ME}}{MSE_{MPS}} \cdot \frac{E(n_{ME})}{E(n_{MPS})},
\end{aligned}$$

تعریف و مورد استفاده قرار گرفته‌اند، که در آن‌ها حجم نمونه و MSE[8] طرح‌ها با اندیس مربوطه نمایش داده شده‌اند. در CeffMW و CeffME به‌دلیل تصادفی بودن حجم نمونه‌ها و اینکه ممکن است حجم نمونه بزرگتر روی بیشتر بودن کارایی تأثیر گذاشته باشد، کارایی توسط هزینه طرح‌ها (در اینجا حجم نمونه‌ها) تعدیل شده است تا تأثیر حجم نمونه از بین رفته و کارایی‌ها به‌صورت عادلانه‌تری با هم مقایسه شوند. بزرگتر از یک بودن این محک‌ها نشان از بهتر بودن روش MPS نسبت به روش مد نظر SRS، MW یا ME دارد و هرچه این مقادیر از یک بزرگتر باشند، این برتری بیشتر است. در Eff حجم نمونه SRS در هر بار شبیه‌سازی مونت کارلو برابر با حجم نمونه استفاده شده توسط MPS در نظر گرفته خواهد شد، که در نتیجه آن مقایسه طرح‌ها عادلانه بوده و نیاز به تعدیل هزینه ندارد. همچنین برای تمامی شبیه‌سازی‌ها نتایج بر اساس ۲۰۰۰۰ تکرار مونت کارلو به‌دست آمده‌اند.

۱.۳ شبیه‌سازی بر اساس توزیع نرمال و پواسون

از توزیع نرمال با میانگین ۱۰۰ و انحراف معیار ۸ و از توزیع پواسون با پارامتر ۱۰، متغیرهای اصلی تولید شدند. در شبیه‌سازی‌ها برای تولید متغیر کمکی بر اساس مدل (۶)، β برابر یک و ε با توجه به همبستگی

[7] Simple random sampling without replacement

[8] Mean square error

مورد نیاز، بازه توزیع یکنواخت بزرگتر یا کوچکتر انتخاب شد. همچنین پارامترهای n_1، n_2 و n_0 در طرح به‌ترتیب معرف امید حجم نمونه روش وزن‌های بیشینه، امید حجم نمونه روش پواسون کلاسیک و حداقل حجم نمونه‌ای که قصد داریم با MPS به آن برسیم بودند که بر اساس آن‌ها نمونه‌هایی تصادفی توسط MPS، MW و ME گرفته شد. جداول ۱ و ۲ نتایج مربوط به کارایی تعدیل شده نسبت به MW و ME به‌ترتیب برای دو توزیع پواسون و نرمال را نشان می‌دهند، که در آن‌ها Ceff در حالت‌های همبستگی زیاد ($\rho = ۰/۹$)، متوسط ($\rho = ۰/۷$) و کم ($\rho = ۰$) به‌دست آمده است. نتایج این شبیه‌سازی را می‌توان به‌صورت زیر فهرست کرد:

- الگوی رفتاری دو طرح MW و ME به‌طور تقریبی در تمامی حالات مشابه یکدیگر بودند بدین معنی که برای همبستگی بالا، یعنی ۰/۹، در بیشتر موارد هر دو روش، عملکرد بهتری نسبت به MPS داشته‌اند و متوسط ستون‌های کارایی تعدیل شده در ردیف آخر، اعداد کمتر از یک را نشان می‌دهند. برای همبستگی متوسط و پایین، یعنی ۰/۷ و ۰/۰، در بیشتر موارد به‌خصوص در نمونه‌های با حجم بالاتر، شاهد بهتر بودن عملکرد MPS نسبت به MW و ME هستیم. همچنین اگر به سطر آخر که میانگین نتایج هر ستون است دقت شود، مشاهده می‌شود که متوسط مقادیر CeffMW و CeffME در این همبستگی بزرگتر از یک هستند که نشان‌دهنده عملکرد بهتر MPS نسبت به هر دو طرح MW و ME است.

- در حالت همبستگی بالا به‌دلیل بهره کامل دو طرح MW و ME از مزیت احتمالات متناسب با اندازه، کارایی بسیار خوبی از خود نشان می‌دهند و بین این دو طرح، MW به‌دلیل در نظر گرفتن شرایط طرح نمونه‌گیری پواسون از نظر کارایی ضعیف‌تر از ME است. بهتر بودن کارایی طرح MPS نسبت به روش MW و ME به‌دلیل استفاده هم‌زمان MPS از نمونه‌گیری تصادفی ساده و احتمالات متناسب با اندازه است. زیرا احتمال متناسب با اندازه برای طرح‌های با همبستگی بالا با متغیر کمکی، و از طرفی دیگر، روش تصادفی ساده برای طرح‌هایی که همبستگی قابل قبولی با متغیر کمکی ندارند، مناسب هستند. به همین دلیل MPS برای داده‌ها با هر مقدار متوسط و پایین، همبستگی کارایی نسبتاً خوبی از خود نشان می‌دهد که البته بهترین وضعیت برای MPS مقدار همبستگی متوسط است، چرا که در این حالت MPS به خوبی از مزیت هر دو طرح متناسب با اندازه و تصادفی ساده استفاده می‌کند.

- از بین سه پارامتر n_0، n_1 و n_2 در طرح‌های نمونه‌گیری، به نظر می‌رسد n_0 بیشترین تاثیر را روی کارایی طرح MPS نسبت به MW و ME دارد و با افزایش آن، کارایی‌های تعدیل شده به‌طور معنی‌داری

جدول ۱: کارایی تعدیل‌شده برای توزیع پواسون

			CeffMW			CeffME		
n_1	n_2	n_0	$\rho = 0$	$\rho = 0/7$	$\rho = 0/9$	$\rho = 0$	$\rho = 0/7$	$\rho = 0/9$
۵	۲	۱	۰٫۷۱	۰٫۷۸	۰٫۶۳	۰٫۷۰	۰٫۷۳	۰٫۵۴
		۲	۱٫۳۲	۱٫۵	۰٫۶۴	۱٫۲۵	۱٫۳۷	۰٫۵۵
		۳	۱٫۹۶	۲٫۶۱	۰٫۵۲	۱٫۹۲	۲٫۴۰	۰٫۴۴
	۴	۳	۱٫۰۴	۱٫۱۶	۰٫۸۵	۱٫۰۳	۱٫۰۴	۰٫۷۵
		۴	۱٫۳۸	۱٫۵۶	۰٫۸۳	۱٫۳۲	۱٫۴۲	۰٫۶۹
		۵	۱٫۹	۲٫۲۶	۰٫۷۱	۱٫۷۵	۲٫۰۸	۰٫۶۰
۷	۴	۳	۰٫۷۶	۱٫۰۶	۰٫۷۱	۰٫۷۴	۰٫۹۳	۰٫۶۰
		۴	۰٫۹۷	۱٫۴۵	۰٫۶۸	۰٫۹۸	۱٫۳۱	۰٫۵۵
		۵	۱٫۳۶	۲٫۰۳	۰٫۶۳	۱٫۳۲	۱٫۷۷	۰٫۵۲
	۶	۵	۰٫۸۹	۱٫۲۱	۰٫۸۲	۰٫۸۸	۱٫۰۵	۰٫۶۹
		۶	۱٫۱۲	۱٫۵۹	۰٫۸	۱٫۰۸	۱٫۴۲	۰٫۶۷
		۷	۱٫۴۳	۲٫۲۴	۰٫۷۶	۱٫۳۴	۱٫۹۵	۰٫۶۲
۹	۶	۵	۰٫۸۳	۱٫۱۳	۰٫۶۷	۰٫۷۷	۰٫۹۹	۰٫۵۲
		۶	۱٫۰۷	۱٫۴۴	۰٫۶۵	۰٫۹۶	۱٫۲۵	۰٫۵۱
		۷	۱٫۴	۱٫۹۶	۰٫۵۹	۱٫۲۹	۱٫۷۱	۰٫۴۶
	۸	۷	۱٫۰۴	۱٫۳	۰٫۸۳	۰٫۹۷	۱٫۱۱	۰٫۶۲
		۸	۱٫۲۳	۱٫۶۷	۰٫۷۶	۱٫۱۳	۱٫۴۲	۰٫۵۹
		۹	۱٫۶۵	۲٫۲۸	۰٫۶۹	۱٫۵۷	۱٫۹۲	۰٫۵۳
میانگین			۱٫۲۲	۱٫۶۲	۰٫۷۱	۱٫۱۷	۱٫۴۴	۰٫۵۸

افزایش یافته‌اند. یکی از دلایل این امر اضافه شدن سهم نمونه تصادفی ساده در MPS است که با توجه به همبستگی پایین موجب بهبود طرح نمونه‌گیری می‌شود.

جدول ۲: کارایی تعدیل‌شده برای توزیع نرمال

			CeffMW			CeffME		
n_1	n_2	n_0	$\rho =$ ۰	$\rho =$ ۰٫۷	$\rho =$ ۰٫۹	$\rho =$ ۰	$\rho =$ ۰٫۷	$\rho =$ ۰٫۹
۵	۲	۱	۰٫۸۳	۰٫۷۶	۰٫۵۳	۰٫۶۹	۰٫۶۵	۰٫۴۷
		۲	۱٫۴۴	۱٫۶۰	۰٫۸۴	۱٫۱۴	۱٫۳۳	۰٫۷۱
		۳	۲٫۷۵	۳٫۵۶	۰٫۸۴	۲٫۲۳	۳٫۰۶	۰٫۷۳
	۴	۳	۱٫۲۵	۱٫۱۱	۰٫۷۵	۱٫۰۰	۰٫۹۳	۰٫۶۳
		۴	۱٫۵۰	۱٫۵۳	۱٫۰۰	۱٫۱۹	۱٫۳۰	۰٫۸۵
		۵	۲٫۲۹	۲٫۵۶	۱٫۰۹	۱٫۸۸	۲٫۱۲	۰٫۹۴
۷	۴	۳	۱٫۱۱	۱٫۰۰	۰٫۶۵	۰٫۸۳	۰٫۸۴	۰٫۵۶
		۴	۱٫۳۳	۱٫۵۳	۰٫۷۹	۱٫۰۲	۱٫۲۷	۰٫۶۸
		۵	۲٫۰۰	۲٫۶۳	۰٫۹۲	۱٫۵۴	۲٫۱۶	۰٫۷۵
	۶	۵	۱٫۲۵	۱٫۱۹	۰٫۷۷	۰٫۹۹	۰٫۹۸	۰٫۶۴
		۶	۱٫۵۷	۱٫۶۰	۰٫۹۴	۱٫۲۱	۱٫۳۰	۰٫۷۷
		۷	۲٫۰۰	۲٫۳۶	۱٫۱۰	۱٫۵۸	۱٫۹۸	۰٫۸۸
۹	۶	۵	۱٫۱۱	۱٫۰۵	۰٫۵۷	۰٫۸۵	۰٫۸۱	۰٫۴۷
		۶	۱٫۳۸	۱٫۴۱	۰٫۷۰	۱٫۰۲	۱٫۰۶	۰٫۵۸
		۷	۱٫۷۵	۲٫۱۳	۰٫۷۸	۱٫۳۳	۱٫۶۴	۰٫۶۳
	۸	۷	۱٫۲۹	۱٫۱۳	۰٫۶۸	۰٫۹۸	۰٫۸۷	۰٫۵۶
		۸	۱٫۵۷	۱٫۵۳	۰٫۸۳	۱٫۲۱	۱٫۱۶	۰٫۶۸
		۹	۲٫۰۰	۲٫۲۱	۰٫۹۱	۱٫۵۰	۱٫۶۸	۰٫۷۶
میانگین			۱٫۵۷	۱٫۷۲	۰٫۸۲	۱٫۲۳	۱٫۴۰	۰٫۶۸

۲.۳ تحلیل داده‌های گیاه بابونه

در این بخش داده‌های مربوط به گیاه دارویی بابونه، تهیه شده در دانشکده کشاورزی دانشگاه صنعتی اصفهان (پناه‌بحق و همکاران، ۲۰۱۶) مورد بررسی قرار می‌گیرد. این گیاه دارای خواص دارویی فراوانی است که از جمله آن‌ها می‌توان به رفع گرفتگی ماهیچه‌های دستگاه گوارش، ضد نفخ بودن، ضد اسهال بودن، درمان سردرد و میگرن، رفع بی‌خوابی و استرس اشاره کرد. محصول اصلی این گیاه، اسانس (عرق) و گلبرگ‌های خشک‌شده

جدول ۳: کارایی MPS برای گیاهان دارویی

$n_۲$	$n_۰$	اسانس $\rho=۰/۷$	$\rho=۰/۵$	$\rho=۰/۳$	وزن خشک $\rho=۰/۹$	$\rho=۰/۸$	$\rho=۰/۶$
۳	۲	۱/۸۵	۱/۳۴	۱/۲۵	۳/۷۴	۱/۶۳	۱/۳۵
	۳	۱/۸۵	۱/۳۹	۱/۲۴	۴/۸۷	۱/۵۴	۱/۲۵
	۴	۱/۵۵	۱/۲۹	۱/۳۰	۵/۲۵	۱/۴۹	۱/۲۶
۶	۴	۱/۹۴	۱/۴۱	۱/۲۸	۳/۶۹	۱/۶۳	۱/۳۹
	۶	۱/۸۳	۱/۳۵	۱/۳۳	۴/۹۶	۱/۵۲	۱/۳۴
	۸	۱/۶۴	۱/۳۳	۱/۳۱	۵/۳۲	۱/۴۲	۱/۳۱
۹	۷	۲/۰۳	۱/۵۲	۱/۳۵	۳/۶۴	۱/۶۱	۱/۳۴
	۹	۱/۸۸	۱/۴۲	۱/۳۸	۴/۵۷	۱/۶۱	۱/۲۹
	۱۱	۱/۷۷	۱/۳۴	۱/۳۵	۵/۳۲	۱/۵۱	۱/۲۶
۱۲	۱۰	۲/۲۱	۱/۵۱	۱/۴۰	۳/۷۷	۱/۵۷	۱/۳۳
	۱۲	۲/۱۶	۱/۵۱	۱/۳۸	۴/۶۱	۱/۶۲	۱/۲۹
	۱۴	۱/۹۹	۱/۴۷	۱/۳۸	۵/۴۵	۱/۴۷	۱/۲۷
میانگین		۱/۸۹	۱/۴۱	۱/۳۳	۴/۶۰	۱/۵۵	۱/۳۱

است. هدف برآورد پارامترهای مجموع (یا میانگین) محصولات تولیدشده توسط این گیاه در یک گلخانه است. دراین‌صورت میزان اسانس و وزن گلبرگ‌های خشک به‌عنوان دو متغیر اصلی در نظر گرفته می‌شوند و برای متغیرهای کمکی می‌توان از اطلاعاتی همچون وزن گلبرگ‌های تر، تعداد ساقه‌های گلدار، تعداد گل‌ها در گیاه،

ارتفاع دم‌گل‌ها و ارتفاع ساقه‌ها استفاده کرد. هر کدام از متغیرهای اصلی و کمکی با هم همبستگی‌های مختلفی دارند. مقادیر همبستگی ۰٫۳ و ۰٫۵ و ۰٫۷ به‌ترتیب همبستگی‌های اسانس و تعداد گل‌ها در گیاه، اسانس و ارتفاع دم گل‌ها و اسانس و وزن تر، و مقادیر همبستگی ۰٫۶ و ۰٫۸ و ۰٫۹ به‌ترتیب مربوط به متغیرهای وزن خشک و تعداد ساقه‌های گل‌دار، وزن خشک و ارتفاع ساقه‌ها، وزن خشک و وزن تر را نشان می‌دهند. متغیرهای کمکی در نظر گرفته شده در این حالت در دسترس و ارزان قیمت (از نظر اندازه‌گیری) هستند. برای ارزیابی MPS در هر بار شبیه‌سازی، نمونه‌هایی تصادفی با پارامترهای مختلف طرح (n_0 و n_2) با استفاده از MPS و SRS با حجم نمونه یکسان گرفته شد. نتایج کارایی برای MPS در جدول ۳ آورده شده‌اند. همان‌طور که ملاحظه می‌شود در این حالت به‌دلیل مقایسه MPS با SRS، تغییرات n_0 موجب تغییرات زیادی در کارایی نشده است. همچنین تمامی کارایی‌ها از یک بزرگ‌تر هستند که نشان‌دهنده عملکرد بهتر طرح MPS در تمامی حالات نسبت به طرح SRS است. طرح MPS به خوبی از اطلاعات متناسب با اندازه استفاده کرده و کارایی آن با افزایش همبستگی افزایش پیدا می‌کند. همان‌گونه که مشاهده می‌شود، MPS در تمامی حالت‌های همبستگی، حتی همبستگی‌های کم نیز مقادیر کارایی بزرگ‌تر از یک را نشان می‌دهد.

۴ پیوست

برای محاسبه احتمال شمول تکی، پیشامد انتخاب شدن عضو k-ام در مرحله اول را با A_1، پیشامد انتخاب شدن عضو k-ام در مرحله دوم را با A_2 و پیشامد اینکه حجم نمونه در مرحله اول بدون در نظر گرفتن عضو مشخص k برابر با b باشد را با B_b نشان داده می‌شود. دراین‌صورت

$$\begin{aligned}
\pi_k^* = P(k \in s) &= P(A_1) + P(A_1^c \cap A_2) = P(A_1) + P(A_1^c)P(A_2|A_1^c) \\
&= \pi_k + (1-\pi_k)\sum_{b=0}^{N} P(A_2 \cap B_b|A_1^c) \\
&= \pi_k + (1-\pi_k)\sum_{b=0}^{N} P(A_2|B_b \cap A_1^c)P(B_b|A_1^c),
\end{aligned}$$

و ازآنجاکه

$$P(A_2|B_b \cap A_1^c) = \frac{n_0 - b}{N - b},$$
$$P(B_b|A_1^c) = \sum_{i=1}^{\binom{N-1}{b}} \prod_{h=1}^{b} \pi_{h_i} \prod_{h=b+1}^{N-1} (1 - \pi_{h_i}),$$

در نتیجه

$$\pi_k^* = \pi_k + (1 - \pi_k) \sum_{b=0}^{n_0 - 1} \frac{n_0 - b}{N - b} \sum_{i=1}^{\binom{N-1}{b}} \prod_{h=1}^{b} \pi_{h_i} \prod_{h=b+1}^{N-1} (1 - \pi_{h_i}).$$

واریانس برآوردگر هارویتز-تامسون به‌صورت

$$\begin{aligned} Var(\hat{t}_\pi) &= V(\sum_{k \in s} \frac{y_k}{\pi_k}) = Var(\sum_{k \in U} \frac{y_k}{\pi_k} I_k) \\ &= \sum_{k \in U} \left(\frac{y_k}{\pi_k}\right)^2 Var(I_k) + \sum_{k \in U} \sum_{\substack{l \in U \\ k \neq l}} \left(\frac{y_k}{\pi_k}\right) \left(\frac{y_l}{\pi_l}\right) \mathrm{Cov}(I_k, I_l) \\ &= \sum_{k \in U} \left(\frac{y_k}{\pi_k}\right)^2 \pi_k (1 - \pi_k) + \sum_{k \in U} \sum_{\substack{l \in U \\ l \neq k}} \left(\frac{y_k}{\pi_k}\right) \left(\frac{y_l}{\pi_l}\right) (\pi_{kl} - \pi_k \pi_l) \\ &= \sum_{k \in U} \sum_{l \in U} \Delta_{kl} \left(\frac{y_k}{\pi_k}\right) \left(\frac{y_l}{\pi_l}\right) \end{aligned}$$

محاسبه می‌شود، که در آن Cov نماد کواریانس است. همچنین برای نشان دادن نااریبی برآوردگر واریانس می‌توان نوشت

$$\begin{aligned} E(\hat{V}(\hat{t}_\pi)) &= E\left(\sum_{k \in s} \sum_{l \in s} (\frac{y_k}{\pi_k})(\frac{y_l}{\pi_l}) \frac{(\pi_k \pi_l - \pi_{kl})}{\pi_{kl}}\right) \\ &= E\left((\sum_{k \in U} \frac{y_k}{\pi_k} I_k)(\sum_{l \in U} \frac{y_l}{\pi_l} I_l)(\frac{\Delta_{kl}}{\pi_{kl}})\right) \\ &= \sum_{k \in U} \sum_{l \in U} (\frac{y_k}{\pi_k})(\frac{y_l}{\pi_l})(\frac{\Delta_{kl}}{\pi_{kl}}) E(I_{kl}) \\ &= \sum_{k \in U} \sum_{l \in U} (\frac{y_k}{\pi_k})(\frac{y_l}{\pi_l}) \Delta_{kl} \end{aligned}$$

و بنابراین با محاسبه احتمالات شمول توام، برهان کامل می‌شود. برای محاسبه احتمال شمول توام، پیشامد انتخاب شدن عضو k-ام در مرحله اول را با A_1، پیشامد انتخاب شدن عضو k-ام در مرحله دوم را با A_2،

پیشامد حضور b واحد در مرحله اول بدون در نظرگرفتن دو عضو k و l را با B_b، پیشامد انتخاب شدن عضو l-ام در مرحله اول را با C_1 و پیشامد انتخاب شدن عضو l-ام در مرحله دوم را با C_2 نشان داده می‌شود. دراین‌صورت

$$\begin{aligned}
\pi^*_{kl} &= P(k \in s, l \in s) = P(A_1 \cap C_1) + P(A_1)P(C_2 \cap C_1^c) + P(C_1)P(A_2 \cap A_1^c) \\
&+ P(C_2 \cap A_2 | C_1^c \cap A_1^c) P(C_1^c \cap A_1^c) \\
&= P(A_1)P(C_1) + P(A_1)P(C_2|C_1^c)P(C_1^c) + P(C_1)P(A_2|A_1^c)P(A_1^c) \\
&+ P(C_2 \cap A_2 | C_1^c \cap A_1^c) P(C_1^c) P(A_1^c)
\end{aligned}$$

و با جایگذاری احتمال‌ها می‌توان نوشت

$$\begin{aligned}
\pi^*_{kl} &= \pi_k \pi_l + \pi_k(1-\pi_l) \sum_{b=0}^{N} P(C_2 \cap B_b | C_1^c) + \pi_l(1-\pi_k) \sum_{b=0}^{N} P(A_2 \cap B_b | A_1^c) \\
&+ (1-\pi_k)(1-\pi_l) \sum_{b=0}^{N} P(A_2 \cap C_2 \cap B_b | A_1^c \cap C_1^c) \\
&= \pi_k \pi_l + \pi_k(1-\pi_l) \sum_{b=0}^{N} P(C_2 | B_b \cap C_1^c) P(B_b | C_1^c) \\
&+ \pi_l(1-\pi_k) \sum_{b=0}^{N} P(A_2 | B_b \cap A_1^c) P(B_b | A_1^c) \\
&+ (1-\pi_k)(1-\pi_l) \sum_{b=0}^{N} P(A_2 \cap C_2 | B_b \cap A_1^c \cap C_1^c) P(B_b | A_1^c \cap C_1^c),
\end{aligned}$$

و ازآنجاکه

$$P(C_2 | B_b \cap C_1^c) = P(A_2 | B_b \cap A_1^c) = \frac{n_0 - b - 1}{N - b - 1},$$
$$P(A_2 \cap C_2 | B_b \cap A_1^c \cap C_1^c) = \frac{n_0 - b(n_0 - b - 1)}{N - b(N - b - 1)},$$

$$
\begin{aligned}
P(B_b|C_1^c) &= P(B_b|A_1^c) \\
&= P(B_b|A_1^c \cap C_1^c) \\
&= \sum_{i=1}^{\binom{N-2}{b}} \prod_{h=1}^{b} \pi_{h_i} \prod_{h=b+1}^{N-2} (1-\pi_{h_i}), \\
P(A_1^c \cap C_1^c) &= P(A_1^c)P(C_1^c),
\end{aligned}
$$

نتیجه می‌شود

$$
\begin{aligned}
\pi_{kl}^* &= \pi_k\pi_l + \pi_k(1-\pi_l)\sum_{b=0}^{n_0-2} \frac{n_0-b-1}{N-b-1} \sum_{i=1}^{\binom{N-2}{b}} \prod_{h=1}^{b} \pi_{h_i} \prod_{h=b+1}^{N-2} (1-\pi_{h_i}) \\
&+ \pi_l(1-\pi_k)\sum_{b=0}^{n_0-2} \frac{n_0-b-1}{N-b-1} \sum_{i=1}^{\binom{N-2}{b}} \prod_{h=1}^{b} \pi_{h_i} \prod_{h=b+1}^{N-2} (1-\pi_{h_i}) \\
&+ (1-\pi_k)(1-\pi_l)\sum_{b=0}^{n_0-2} \frac{n_0-b(n_0-b-1)}{N-b(N-b-1)} \sum_{i=1}^{\binom{N-2}{b}} \prod_{h=1}^{b} \pi_{h_i} \prod_{h=b+1}^{N-2} (1-\pi_{h_i}).
\end{aligned}
$$

بحث و نتیجه‌گیری

در این مقاله یک طرح نمونه‌گیری پواسون اصلاح‌شده ارائه شد که با حفظ استقلال نسبی انتخاب اعضا، تضمین‌کننده حداقلی از حجم نمونه است. این طرح نمونه‌گیری برای نمونه‌گیری متناسب با اندازه، حتی در حالت همبستگی کم با متغیر کمکی نیز مناسب است. شبیه‌سازی‌ها تایید می‌کنند که این طرح، نسبت به طرح تصادفی ساده و طرح سنتی نمونه‌گیری پواسون بسیار کاراتر و نسبت به دیگر طرح‌های پواسون اصلاح‌شده (MW) و متناسب با اندازه (ME) در حالت همبستگی متوسط و کم، کاراتر است.

تقدیر و تشکر

نویسندگان بر خود لازم می‌دانند از داوران و ویراستار محترم که راهنمایی‌هایشان منجر به ارتقای سطح کیفی مقاله شده است، تشکر کنند.

مراجع

[1] Andersen, I. T., Hahn, U., and Vedel Jensen, E. B. (2015), Optimal PPS Sampling with Vanishing Auxiliary Variables with Applications in Microscopy, *Scandinavian Journal of Statistics*, **42**, 1136-1148.

[2] Bondesson, L., and Thorburn, D. (2008), A List Sequential Sampling Method Suitable for Real-Time Sampling, *Scandinavian Journal of Statistics*, **35**, 466-483.

[3] Brewer, K. R. W., Early, L. J., and Hanif, M. (1980), Poisson, Modified Poisson and Collocated Sampling, *Journal of Statistical Planning and Inference*, **10**, 15-30.

[4] Brewer, K. R. W., and Hanif, M. (1983), *An Introduction to Sampling with Unequal Probabilities*, In: Sampling With Unequal Probabilities, Lecture Notes in Statistics, Vol. 15, Springer, New York.

[5] Chen, X., Dempster, A., and Liu, J. (1994), Weighted Finite Population Sampling to Maximize Entropy, *Biometrika*, **81**, 457-469.

[6] Deville, J. C., and Tille, Y. (1998), Unequal Probability Sampling without Replacement Through a Splitting Method, *Biometrika*, **85**, 89-113.

[7] Ghosh, D., and Vogt, A. (1998), Rectification of Sample Size in Bernoulli and Poisson Sampling, *Proceedings of the American Statistical Association, Survey Research Methods Section*, 448-450.

[8] Grafstrom, A. (2005), Comparisons of Methods for Generating Conditional Poisson Samples and Samford Samples, *Master Thesis*, *Umea University*.

[9] Grafstrom, A. (2010), On a Generalization of Poisson Sampling, *Journal of Statistical Planning and Inference*, **140**, 982-991.

[10] Grafstrom, A. (2010), Entropy of Unequal Probability Sampling Designs, *Statistical Methodology*, **7**, 84-97.

[11] Grafstrom, A. (2012), Spatially Correlated Poisson Sampling, *Journal of Statistical Planning and Inference*, **142**, 139-147.

[12] Grafstrom, A. (2015), Coordination of Conditional Poisson Samples, *Journal of Official Statistics*, **31**, 649-672.

[13] Hajek, J. (1964), Asymptotic Theory of Rejective Sampling with Varying Probabilities from a Finite Population, *The Annals of Mathematical Statistics*, **35**(4), 1491-1523.

[14] 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.

[15] Narain, R. (1951), On Sampling without Replacement with Varying Probabilities, *Journal of the Indian Society of Agricultural Statistics*, **3**, 169-175.

[16] Ogus, J., and Clark, D. (1971), *The Annual Survey of Manufactures: A Report on Methodology*, U.S. Department of Commerce, Bureau of the Census, Washington DC.

[17] Panahbehagh, B., Bruggemann, B., Parvardeh, A., Salehi, M., and Sabzalian, M. R. (2017), An Unbalanced Ranked-Set Sampling Method to Get More Than One Sample from Each Set, *Journal of Survey Statistics and Methodology*, smx026. https://doi.org/10.1093/jssam/smx026

[18] Sarndal, C. E., Swensson, B., and Wretman, J. (2003), *Model Assisted Survey Sampling*, Springer, New York.

[19] Tillé, Y. (2006), Sampling Algorithms, *Springer Series in Statistics*, Springer Science + Business Media Inc., New York.

[20] Tillé, Y., and Wilhelm, M. (2017), Probability Sampling Designs: Principles for Choice of Design and Balancing, *Statist. Sci.,* **32**, 176-189.

An Efficient Poisson Sampling Design in Proportional to Size Situation

Afshin, P., Panahbehagh, B., Sanatpour, A. H.

Department of Mathematics, Kharazmi University, Tehran, Iran.

Abstract: We introduce a modified Poisson sampling, with a fixed lower bound of sample size. The design is a combination of simple random sampling and Poisson sampling. Simple random sampling is used to compensate for the lack of sample size from remaining elements in the finite population, after execution of a Poisson sampling. At the first stage, the units are sampled independently with given inclusion probabilities. But in the second stage, inclusion probabilities are dependent to each other. Because it is important to know, which of the elements are selected in the first stage and which of them are remained. Some advantages of our design are: simple performance, controlling sample size, ability to perform the method of probability proportional to size. The simulations show that the design can dominate its rival design in probability proportional to size sampling.

Keywords: Modified Poisson Sampling, Efficiency, Probability Proportional to Size.

Mathematics Subject Classification (2010): 62D05, 62N02, 62L05.