مقایسه‌ی طرح‌های πps با معیار آنتروپی و فاصله‌ی هلینگر

فهیمه مسیحی بیدگلی، الهام همایونی
دانشکده ریاضی، دانشگاه صنعتی اصفهان

چکیده

هدف اصلی در نمونه‌گیری، انتخاب یک نمونه از جامعه به‌منظور برآورد بعضی از پارامترهای نامعلوم جامعه است. هنگامی که احتمال انتخاب واحدها برای مشمول شدن در نمونه یکسان نباشد، نمونه‌گیری با احتمال نابرابر مطرح می‌شود. معمولاً احتمال‌های شمول را متناسب با متغیر کمکی که برای همه‌ی واحدها در جامعه معلوم است، انتخاب می‌کنند که در صورت بدون جایگذاری بودن روش نمونه‌گیری، طرح πps حاصل می‌شود. وقتی نمونه‌گیری با احتمال نابرابر قابل اجرا باشد، برآوردهای بهتری نسبت به نمونه‌گیری با احتمال برابر وجود دارد. برای نمونه‌گیری با احتمال نابرابر، طرح‌های نمونه‌گیری مختلفی می‌توان به‌کار برد. انتخاب طرح نمونه‌گیری دارای اهمیت است. زیرا توسط طرح خصوصیات برآوردگری که استفاده می‌گردد، تعیین می‌شود. اجرای ساده و کارای یک طرح به‌منظور استفاده‌ی عملی توسط آماردانان مورد نظر است. همچنین یک طرح نمونه‌گیری باید یک سطح بالایی از تصادفی بودن داشته باشد. یک اندازه از میزان تصادفی بودن، آنتروپی است. معیار دیگر مقایسه‌ی طرح‌های نمونه‌گیری، محاسبه‌ی فاصله‌ی طرح‌هاست که یکی از این اندازه‌ها فاصله‌ی هلینگر است. در این مقاله با محاسبه‌ی آنتروپی و فاصله‌ی هلینگر چندین طرح معروف πps، طرح‌های برتر و مشابه مشخص شده‌اند.

واژه‌های کلیدی: آنتروپی، طرح نمونه‌گیری، فاصله‌ی هلینگر، طرح‌های πps.

۱ مقدمه

یکی از روش‌های نمونه‌گیری با احتمال نابرابر، نمونه‌گیری متناسب با اندازه و بدون جایگذاری است که با نماد πps نمایش داده می‌شود. در این روش هر یک از N واحد در جامعه با احتمال شمول $\pi_i(i = ۱, \ldots, N)$ وارد نمونه می‌شود. معمولاً π_iها طوری انتخاب می‌شوند که

$\sum_{i=1}^{N} \pi_i = n$ که در آن n اندازه‌ی نمونه است. اگر هر واحدی که انتخاب می‌شود، مستقل از سایر واحدها با احتمال π_i انتخاب شود و نمونه‌ی ممکن بدون در نظر گرفتن اندازه‌ی آن پذیرفته شود، طرح نمونه‌گیری پواسون[۱] حاصل می‌شود. برای به‌دست آمدن اندازه‌ی نمونه‌ی n می‌توان طرح نمونه‌گیری فوق را به طریق‌های مختلفی اصلاح کرد. این روش‌ها عبارت‌اند از: نمونه‌گیری پواسون شرطی (cp)[۲] (هاجک [۶، ۷])، نمونه‌گیری سامفورد ($Sampf$)[۳] (سامفورد [۱۲])، نمونه‌گیری پارتو (par)[۴] (رسن [۱۰، ۱۱])، روش بریور[۵] (بریور [۲]) و روش نمونه‌گیری سیستماتیک (مادو [۸]). هر یک از این طرح‌ها دارای تابع احتمال مشخص بوده و با استفاده از توزیع بردار طرح حاصل می‌شوند. روش خرد کردن، یک کلاس کلی از روش‌های نمونه‌گیری با احتمال نابرابر و بدون جایگذاری است. این روش توسط دویل و تیل [۵] معرفی و در تیل [۱۴] مورد بحث قرار گرفت. از جمله روشهای نمونه‌گیری به این شیوه، طرح خرد کردن به نمونه‌گیری تصادفی ساده و روش محوری است. سؤالی که مطرح می‌شود این است که چگونه می‌توان طرح‌های مختلف را با یکدیگر مقایسه کرد. ممکن است در یک جامعه‌ی خاص، واریانس برآوردگر جمع جامعه برای یک طرح، کوچک باشد و در جامعه‌ی دیگر همان طرح نتیجه‌ی دیگری داشته باشد. یکی از معیارهای خوب بودن طرح، تصادفی بودن آن است. یک اندازه از تصادفی بودن با آنتروپی شانون نشان داده می‌شود که اولین بار توسط شانون [۱۳] معرفی شد. آنتروپی طرح نمونه‌گیری به‌صورت زیر تعریف می‌شود.

$$H = -\sum_{\mathbf{x}\in\mathcal{Q}} p(\mathbf{x})\log p(\mathbf{x}) = -E_p\ [\ \log(p(\mathbf{x}))\ ],$$

که در آن $\mathcal{Q}$ مجموعه‌ای از نمونه‌های ممکن و به تکیه‌گاه طرح معروف است.

طرح‌های با آنتروپی بالا، استوار هستند. همچنین در طرح‌های با آنتروپی بالا، برآوردگر هارویتز-تامپسون بدون نیاز به محاسبه‌ی احتمال‌های شمول مرتبه‌ی دوم به‌خوبی برآورد می‌کند و سرعت همگرایی بالایی به توزیع نرمال دارد. اولین بار هاجک [۷] نشان داد که طرح پواسون شرطی تعدیل شده دارای بیشترین میزان آنتروپی است.

به‌منظور مقایسه‌ی طرح‌های مختلف می‌توان از معیار فاصله‌ی هلینگر بین آن‌ها استفاده کرد که این فاصله به‌صورت زیر تعریف می‌شود.

$$D_H^2(p_1,p_2) = \frac{1}{2}\sum_{\mathbf{x}\in\mathcal{Q}}(\sqrt{p_1(\mathbf{x})} - \sqrt{p_2(\mathbf{x})})^2,$$

که در آن $\mathcal{Q} = \mathcal{Q}_1 \cup \mathcal{Q}_2$ و $\mathcal{Q}_1$ و $\mathcal{Q}_2$ به ترتیب تکیه‌گاه‌های p_1 و p_2 هستند. $D_H(p_1,p_2)$ عددی در بازه‌ی $[0,1]$ است. هر چه این فاصله به صفر نزدیک‌تر باشد، بیانگر نزدیکی بیشتر دو تابع احتمال p_1 و p_2 خواهد بود.

ابتدا در بخش ۲ به معرفی بعضی از طرح‌های πps و شیوه‌ی اجرایی آن‌ها پرداخته می‌شود و در بخش ۳ تابع احتمال این طرح‌ها معرفی می‌شود. در بخش ۴ با بررسی دو جامعه‌ی معروف، آنتروپی و فاصله‌ی هلینگر این طرح‌ها محاسبه شده و در انتها نتایج بیان می‌شود.

---

[۱] Poisson sampling design

[۲] Conditional Poisson Sampling

[۳] Sampford sampling

[۴] Pareto sampling

[۵] Brewer's method

۲ معرفی بعضی از طرح‌های πps

در این بخش برخی از مهم ترین طرح های πps را معرفی و تشریح می‌کنیم.

۱.۲ نمونه‌گیری پواسون شرطی

نمونه‌گیری پواسون شرطی توسط هاجک [۶، ۷] و مورد بررسی قرار گرفت. به‌طور متوالی نمونه‌های پواسون تولید می‌شوند و تنها در صورتی که اندازه‌ی نمونه n باشد، نمونه‌ی مورد نظر به‌عنوان یک نمونه‌ی پواسون شرطی پذیرفته می‌شود. از آنجا که همه‌ی نمونه‌ها پذیرفته نمی‌شوند، این شیوه روی احتمال‌های شمول واقعی تأثیر می‌گذارد. در صورتی‌که احتمال‌های شمول از پیش تعیین‌شده برای طرح پواسون به‌عنوان پارامترهای نمونه‌گیری طرح پواسون شرطی با $\mathrm{p}_i\ (i=۱,۲,\ldots,N)$ نمایش داده شود، در این صورت هر واحد i به‌طور مستقل از بقیه‌ی واحدها با احتمال p_i انتخاب می‌شود. این طرح دارای اجراهای دیگر نیز است که از کارایی و سرعت بیشتری برخورداراند.

۲.۲ نمونه‌گیری پواسون شرطی تعدیل شده

در نمونه‌گیری پواسون شرطی، احتمال‌های شمول واقعی با پارامترهای نمونه‌گیری برابر نیست. می‌توان با فرض p_i ها به‌عنوان احتمال‌های شمول هدف π_i، پارامترهای طرح CP را به‌منظور ایجاد احتمال‌های شمول هدف اصلاح کرد. در صورت استفاده از پارامترهای تعدیل‌شده در طرح پواسون شرطی، این طرح را پواسون شرطی تعدیل‌شده(ACP)[۶] می‌نامند. روش‌های بسیاری برای اصلاح پارامترهای طرح CP توسط برخی از محققین پیشنهاد شده است. از جمله روش‌های دقیق، روش تکرار است که به دو صورت شیوه‌ی تکرار آیرس [۱] و شیوه‌ی تکرار چن و همکاران [۴] اجرا می‌شود.

۳.۲ نمونه‌گیری سامفورد

نمونه‌گیری سامفورد، توسط سامفورد [۱۲] معرفی شد. این روش یک طرح دقیق است و احتمال‌های شمول از پیش تعیین‌شده‌ی $\pi_i\ (i=۱,۲,\ldots,N)$ را ایجاد می‌کند. به‌عبارت دیگر در این طرح، احتمال‌های شمول واقعی با احتمال‌های شمول هدف برابر خواهند بود. اجرای اصلی طرح فوق بدین شرح است.
واحد اول با احتمال $\frac{\pi_i}{n}\ (i=۱,۲,\ldots,N)$ استخراج می‌شود. با بازگشت عنصر انتخاب‌شده به جامعه $n-۱$ واحد دیگر با جایگذاری و با احتمال $\mathrm{p}'_i \propto \frac{\pi_i}{۱-\pi_i}$ استخراج می‌شوند، به‌طوری‌که $\sum_{i=۱}^{N}\mathrm{p}'_i=۱$. اگر تمام n واحد استخراجی متفاوت باشند، نمونه پذیرفته خواهد شد. در غیر این صورت، فرآیند از ابتدا آغاز می‌شود. این طرح نیز دارای اجراهای دیگری است که دارای کارایی و سرعت بالاتری هستند.

۴.۲ روش بریور

روش بریور[۷] ابتدا توسط بریور [۲] برای مورد خاص $n=۲$ پیشنهاد شد و سپس وی در سال ۱۹۷۵ این روش را برای هر اندازه‌ی نمونه گسترش داد ([۳]). این طرح نیز دقیق بوده و قادر است احتمال‌های از پیش تعیین‌شده‌ی π_i را ایجاد کند. نمونه در n گام انتخاب می‌شود. احتمال‌های هر استخراج به‌صورت

[۶] Adjusted Conditional Poisson design

[۷] Brewer's method

جداگانه تعریف می‌شوند. اگر برای هر واحد، اندازه‌ی $(i=1,\dots,N)$ $\mathrm{p}_i=\frac{\pi_i}{n}$ تعریف شود، آن‌گاه واضح است که $\sum_{i=1}^N \mathrm{p}_i=1$. انتخاب واحدها در $(n-r+1)$-اُمین استخراج (r-اُمین استخراج باقیمانده) از میان واحدهایی که تاکنون انتخاب نشدند، با احتمال متناسب با $\frac{\mathrm{p}_i(1-\mathrm{p}_i)}{1-r\mathrm{p}_i}$ صورت می‌گیرد.

۵.۲ نمونه‌گیری پارتو

طرح نمونه‌گیری پارتو توسط رسن [۱۰، ۱۱] معرفی شد. رسن نشان داد که در کلاس نمونه‌گیری ترتیبی[8]، نمونه‌گیری پارتو دارای حداقل واریانس برای برآوردگر مجموع جامعه است. طرح نمونه‌گیری ترتیبی با در نظر گرفتن متغیرهای تصادفی رتبه‌ای مستقل از هم $Q_i\ (i=1,\dots,N)$ با تابع توزیع $F_i(x)$ و تابع چگالی $f_i(x)$ روی $x\in[0,\infty)$ تعریف می‌شود. واحدهای متناظر با n تا از کوچکترین Q-مقدارها یک نمونه را تشکیل می‌دهد. یک زیرکلاس مهم با فرض آن که همه‌ی F_i ها متعلق به خانواده‌ای با مقیاس یکسان و با تابع توزیع $F_i(x)=\frac{x\theta_i}{1+x\theta_i}$ باشند، حاصل می‌شود که مربوط به شیوه‌ی نمونه‌گیری پارتو است. می‌توان به‌جای پارامتر θ_i از مجموعه پارامتر دیگر به نام $\lambda_i=F_i(1)=\frac{\theta_i}{1+\theta_i}$، به‌ازای هر واحد جامعه استفاده کرد و از آن برای تخمین احتمال‌های شمول استفاده نمود، زیرا $\sum_{i=1}^N\lambda_i=n$ است. اگر برای هر $\lambda_i\ (i=1,2,\dots,N)$ به‌عنوان پارامتر طرح نمونه‌گیری پارتو باشد، برای داشتن یک طرح πps، λ_iها مشابه احتمال‌های شمول هدف، متناسب با اندازه‌ی صفت کمکی انتخاب می‌شوند. در این نمونه‌گیری، ابتدا اعداد تصادفی مستقل $U_1,U_2,\dots,U_N$ از $U(0,1)$ تولید شده، سپس متغیرهای رتبه‌ای پارتو به‌صورت زیر محاسبه می‌شوند.

$$Q_i=\frac{U_i/(1-U_i)}{\lambda_i/(1-\lambda_i)},\qquad i=1,2,\dots N.$$

n واحد متناظر با کوچکترین Q-مقدارها به‌عنوان یک نمونه‌ی پارتو انتخاب می‌شوند. طرح پارتو یک طرح دقیق نیست، زیرا احتمال شمول واقعی برای واحد i با λ_i برابر نیست.

۶.۲ نمونه‌گیری پارتو تعدیل‌شده

در طرح نمونه‌گیری پارتو، همانند طرح نمونه‌گیری پواسون شرطی، می‌توان پارامترهای طرح نمونه‌گیری را اصلاح کرد تا احتمال‌های شمول هدف حاصل شود که در این صورت طرح پارتو تعدیل شده[9] به‌وجود می‌آید. دو شیوه‌ی تکرار آیرس [۱] و چن و همکاران [۴] در این طرح نیز استفاده می‌شود.

۷.۲ نمونه‌گیری سیستماتیک

برای اولین بار نمونه‌گیری سیستماتیک با احتمال متناسب با اندازه[10] توسط مادو [۸] پیشنهاد شد. این طرح به علت سادگی در اجرا به‌طور گسترده در بررسی‌های نمونه‌ای مورد استفاده قرار می‌گیرد. در این روش $V_k\ (k=0,\dots,N)$ به‌صورت زیر تعریف می‌شود.

$$V_k=\sum_{l=1}^k \pi_l.$$

واضح است که $V_0=0$ و $V_N=n$. نمونه‌گیری سیستماتیک به روش متناسب با اندازه با تولید یک عدد تصادفی u از متغیر تصادفی یکنواخت $U[0,1]$ اجرا می‌شود. نمونه‌ی x شامل تمام واحدهای i خواهد بود که

---

[8] Order Sampling

[9] Adjusted Pareto sampling design

[10] Systematic πps sampling

در نامساوی زیر صدق کنند.

$$V_{i-1} < k+u \le V_i, \qquad (k=0,1,\dots,n-1).$$

این طرح بستگی به ترتیب واحدهای جامعه داشته و به طرح نمونه‌گیری سیستماتیک ترتیبی[11] نیز معروف است. پی و همکاران [۹] نشان دادند که طرح نمونه‌گیری سیستماتیک، یک طرح با مینیمم تکیه‌گاه است.

۸.۲ طرح خرد کردن به نمونه‌گیری تصادفی ساده ($SSRS$)

طرح خرد کردن به نمونه‌گیری تصادفی ساده($SSRS$)[12] توسط دویل و تیل [۵] معرفی شد. این طرح یک مورد خاص از نمونه‌گیری خرد کردن است. در این طرح بردار احتمال شمول به دو بردار خرد می‌شود. یک بردار شامل احتمال‌های شمول برابر و بردار دیگر بر اساس قاعده‌ی خاصی ایجاد می‌شود. یکی از این دو بردار به‌صورت تصادفی انتخاب می‌شود. اگر بردار با احتمال‌های شمول برابر انتخاب شد، یک نمونه‌گیری تصادفی ساده صورت می‌گیرد و در صورت انتخاب بردار دیگر، دوباره عمل خرد کردن تکرار می‌شود. گام نخست این شیوه به‌صورت زیر توصیف می‌شود:

اگر بردار π بردار شروع باشد، مقدار α عبارت است از:

$$\alpha=\min\left(\frac{N}{n}\,\pi_{(1)}\ ,\ \frac{N}{N-n}\,(1-\pi_{(N)})\right),$$

که در آن $\pi_{(N)},\dots,\pi_{(k)},\dots,\pi_{(1)}$ احتمال‌های شمول ترتیبی هستند.

برای $k\in U$ مقادیر $\pi_k^{(1)}$ و $\pi_k^{(2)}$ به‌صورت زیر محاسبه می‌شود:

$$\pi_k^{(1)}=\frac{n}{N}, \qquad \pi_k^{(2)}=\frac{\pi_k-\alpha(\frac{n}{N})}{1-\alpha}.$$

بنابراین بردارهای جدید عبارت‌اند از:

$$\pi^{(1)}=(\frac{n}{N},\frac{n}{N},\dots,\frac{n}{N}),$$

$$\pi^{(2)}=(\pi-\alpha\pi^{(1)})/(1-\alpha).$$

با احتمال α، $\pi^{(1)}$ به‌عنوان بردار π جدید انتخاب می‌شود و با احتمال $1-\alpha$، بردار $\pi^{(2)}$ انتخاب خواهد شد. در صورت انتخاب $\pi^{(1)}$ این شیوه با یک نمونه‌ی تصادفی ساده پایان می‌پذیرد و اگر $\pi^{(2)}$ انتخاب شود، مختصاتی که به ۰ و یا ۱ تبدیل شده‌اند، از بردار حذف شده و در گام بعدی مسأله‌ی نمونه‌گیری به انتخاب یک نمونه با اندازه‌ی n یا $n-1$ از جامعه‌ای با $N-1$ واحد تبدیل می‌شود. نمونه حداکثر در $N-1$ گام حاصل می‌شود.

۹.۲ روش محوری (Piv)

روش محوری[13] توسط دویل و تیل [۵] معرفی شد. این روش یک مورد خاص از روش خرد کردن است و برای انتخاب نمونه‌ی πps یک روش ساده است. الگوریتم روش محوری حداکثر N گام دارد. در هر گام احتمال‌های شمول برای دو واحد جامعه به‌هنگام می‌شود. اگر این دو واحد به‌طور تصادفی انتخاب شوند، آنتروپی این طرح بالا خواهد بود.

اگر در گام t دو واحد انتخابی در جامعه با i و j نشان داده شوند، طوری‌که $0<\pi_i(t)<1$ و $0<\pi_j(t)<1$، آن‌گاه اگر $\pi_i(t)+\pi_j(t)>1$، داریم:

---

[11] Order systematic sampling

[12] Splitting into Simple Random Sampling

[13] Pivotal method

$$\alpha(t)=\frac{1-\pi_j(t)}{2-\pi_i(t)-\pi_j(t)},$$

$$\pi_k^{(1)}(t)=\begin{cases}\pi_k(t), & k\in\mathcal{U}\backslash\{i,j\},\\ 1, & k=i,\\ \pi_i(t)+\pi_j(t)-1, & k=j,\end{cases}$$

$$\pi_k^{(2)}(t)=\begin{cases}\pi_k(t), & k\in\mathcal{U}\backslash\{i,j\},\\ \pi_i(t)+\pi_j(t)-1, & k=i,\\ 1, & k=j.\end{cases}$$

و در صورتی که $\pi_i(t)+\pi_j(t)<1$، آنگاه

$$\alpha(t)=\frac{\pi_i(t)}{\pi_i(t)+\pi_j(t)},$$

$$\pi_k^{(1)}(t)=\begin{cases}\pi_k(t), & k\in\mathcal{U}\backslash\{i,j\},\\ \pi_i(t)+\pi_j(t), & k=i,\\ 0, & k=j,\end{cases}$$

$$\pi_k^{(2)}(t)=\begin{cases}\pi_k(t), & k\in\mathcal{U}\backslash\{i,j\},\\ 0, & k=i,\\ \pi_i(t)+\pi_j(t), & k=j.\end{cases}$$

این روش تا تبدیل همه‌ی مختصات π به 0 و یا 1 ادامه می‌یابد.

۳ تابع احتمال طرح‌های πps

با داشتن تابع احتمال می‌توان ویژگی‌های بیشتری از طرح همچون آنتروپی طرح را نیز مورد مطالعه قرار داد. یک روش برای محاسبه‌ی تابع احتمال، توزیع بردار طرح، $Pr(\mathbf{I}=\mathbf{x})$، بوده که در آن $\mathbf{x}=(x_1,x_2,\dots,x_N)$ یک تحقق از بردار طرح $\mathbf{I}$ است. یک رویکرد کلی دیگر برای یافتن تابع احتمال، استفاده از الگوریتم نمونه‌گیری است. الگوریتم نمونه‌گیری شیوه‌ای برای تولید یک نمونه‌ی تصادفی است.

فرض می‌شود $\pi=(\pi_1,\dots,\pi_N)$ احتمال‌های شمول هدف باشد. تابع احتمال طرح پواسون عبارت است از:

$$p_{poisson}(\mathbf{x})=\prod_{i=1}^{N}\pi_i^{x_i}(1-\pi_i)^{1-x_i}.$$

طرح پواسون شرطی نیز به‌علت رد بعضی از نمونه‌های حاصله تنها یک ثابت برای از بین بردن قسمت رد دارد.

$$p_{CP}(\mathbf{x})=C_{CP}\prod_{i=1}^{N}p_i^{x_i}(1-p_i)^{1-x_i},\qquad \mathbf{x}\in\ \mathcal{S}_n,$$

که در آن $p_i(i=1,\dots,N)$ پارامتر نمونه‌گیری برای واحد i–اُم و ثابت C_{CP} با شرط نرمال‌سازی $\sum_{\mathbf{x};|\mathbf{x}|=n}p_{CP}(\mathbf{x})=1$ به‌دست می‌آید. در صورتی که $d=\sum_{i=1}^{N}p_i(1-p_i)$ به اندازه‌ی کافی بزرگ باشد، می‌توان نشان داد $C_{CP}\approx\sqrt{2\pi d}$ است.

طرح پواسون شرطی تعدیل شده نیز با اصلاح پارامترهای نمونه‌گیری به‌صورت زیر خواهد بود.

$$p_{Acp}(\mathbf{x})=C_{Acp}\prod_{i=1}^{N}(Ap)_i^{x_i}(1-(Ap)_i)^{1-x_i},\ |\mathbf{x}|=n,$$

که در آن $(Ap)_i$ پارامترهای تعدیل‌شده‌اند.

در نمونه‌گیری سامفورد تابع احتمال عبارت است از:

$$p_S(\mathbf{x})=C_S\prod_{i=1}^{N}\pi_i^{x_i}(1-\pi_i)^{1-x_i}\sum_{j=1}^{N}x_j(1-\pi_j),$$

که در آن ثابت C_S با استفاده از شرط نرمال‌سازی $\sum_{\mathbf{x}:\ |\mathbf{x}|=n}p_S(\mathbf{x})=1$، قابل محاسبه است. برای مقادیر بزرگ $d=\sum_{i=1}^{N}\pi_i(1-\pi_i)$، $C_S\approx\sqrt{\frac{2\pi}{d}}$ است.

در نمونه‌گیری پارتو با در نظر گرفتن $\lambda_i(i=1,\dots,N)$ به‌عنوان پارامترهای نمونه‌گیری، تابع احتمال به‌صورت زیر است.

$$p_{pareto}(\mathbf{x})=\prod_{i=1}^{N}\lambda_i^{x_i}(1-\lambda_i)^{1-x_i}\times\sum_{j=1}^{N}c_jx_j,\ |\mathbf{x}|=n,$$

که در آن $c_j=\int_0^{\infty}t^{n-1}\prod_{i=1}^{N}\frac{1+\theta_i}{1+\theta_it}\times\frac{1}{1+\theta_jt}\,dt$، و تابع احتمال طرح پارتو تعدیل‌شده نیز به‌صورت فوق است، تنها به‌جای پارامتر نمونه‌گیری λ_i پارامتر اصلاح شده‌ی $(A\lambda)_i$ قرار می‌گیرد.

در روش بریور با در نظر گرفتن $\mathrm{p}_k^{(i)}$ به‌صورت زیر

$$\mathrm{p}_k^{(i)}=\left[\sum_{j=1}^{N}(1-x_j)\frac{\pi_j(n-\sum_{l\in\mathcal{U}}x_l\pi_l-\pi_j)}{n-\sum_{l\in\mathcal{U}}x_l\pi_l-\pi_j\{n-(i-1)\}}\right]^{-1}\times(1-x_k)\frac{\pi_k(n-\sum_{l\in\mathcal{U}}x_l\pi_l-\pi_k)}{n-\sum_{l\in\mathcal{U}}x_l\pi_l-\pi_k\{n-(i-1)\}},$$

تابع احتمال طرح بریور

$$p_{Brewer}(\mathbf{x})=\sum_{j=1}^{n!}\prod_{i=1}^{n}\mathrm{p}_{k_{ji}}^{(i)},$$

خواهد بود. در نمونه‌گیری سیستماتیک نیز تابع احتمال $p_{syst}(s)=\max_{u:\ u\Rightarrow s}(u)-\min_{u:\ u\Rightarrow s}(u)$ است که $u\Rightarrow s$ عدد تصادفی از توزیع یکنواخت $U(0,1)$ است که به نمونه‌ی s منجر می‌شود. برای یافتن تابع احتمال طرح خرد کردن به نمونه‌گیری تصادفی ساده، فرض کنید α_i احتمال انتخاب $\pi^{(1)}$ در خرد کردن i-اُم باشد و تعداد خروجی‌ها در درخت خرد کننده[14] برابر k باشد. بنابراین بیشینه‌ی تعداد خردکردن‌ها $k-1$ است. اگر خروجی‌ها با $E_k,\dots,E_1,E_2$ نشان داده شود، تابع احتمال به‌صورت زیر خواهد بود:

$$p(\mathbf{x})=\sum_{i=1}^{k}Pr(\mathbf{x}|E_i)Pr(E_i),$$

وقتی

$$Pr(\mathbf{x}|E_i)=\begin{cases}\dfrac{1}{\binom{N_i}{n_i}}, & \pi_k\in\{0,1\},\\ 0, & o.w.\end{cases}$$

π_k مؤلفه‌های بردار در E_i است، $N_i=\#\{\pi_k\notin\{0,1\}\}$ و π_k $\sum_{k:\ \pi_k\notin\{0,1\}}$ $n_i=$، برای $i=1,\dots,k$.

$Pr(E_i)$ نیز با حاصل‌ضرب احتمال‌هایی که منجر به E_i می‌شود، قابل محاسبه خواهد بود.

$$Pr(E_i)=\begin{cases}\alpha_1, & i=1,\\ \alpha_i\prod_{j=1}^{i-1}(1-\alpha_j), & i=2,3,\dots,k-1,\\ \prod_{j=1}^{i-1}(1-\alpha_j), & i=k.\end{cases}$$

یافتن تابع احتمال برای روش محوری در حالت کلی امکان‌پذیر نیست. مشکل اصلی وجود مسیرهای زیاد در منجر شدن به نمونه‌های یکسان است. اگر در هر N گام، دو واحد به‌صورت تصادفی انتخاب شوند، مجموع کل مسیرهای ایجادشده عبارت است از:

$$\binom{N}{2}\binom{N-1}{2}\dots\binom{2}{2}2^{N-1}=N!(N-1)!.$$

احتمال نمونه‌ی x جمع مسیرهایی است که منجر به این نمونه می‌شود. این مسیرها قابل تشخیص نخواهد بود. برای یافتن احتمال نمونه‌های طرح محوری از شبیه‌سازی استفاده می‌شود.

[14]Splitting tree

۴ مثال‌ها

مثال **۱.۴.** *جامعه‌ی معروف به ترات-باندسون-میستر با اندازه‌ی جامعه‌ی ۶ = N و اندازه‌ی نمونه‌ی ۳ = n و احتمال‌های شمول از پیش تعیین‌شده‌ی زیر را در نظر بگیرید:*

$$\pi = (\frac{1}{3}, \frac{1}{3}, \frac{1}{3}, \frac{2}{3}, \frac{2}{3}, \frac{2}{3}).$$

آنتروپی طرح‌های مورد بررسی را از طریق تعریف و با بررسی تمام نمونه‌های ممکن طرح نمونه‌گیری محاسبه کرده و نتایج در جدول ۱ آورده شده است. فاصله‌ی هلینگر طرح‌های مذکور نیز در جدول ۲ ارائه شده است. در جدول ۱ مشاهده می‌شود، نمونه‌گیری پواسون شرطی تعدیل شده دارای حداکثر میزان آنتروپی است که هاجک [۷] این مطلب را به اثبات رسانده است. چهار طرح نمونه‌گیری برتر نمونه‌گیری پواسون شرطی تعدیل‌شده، پارتو تعدیل‌شده، بریور و سامفورد است که طرح‌هایی با حداکثر میزان آنتروپی هستند. در این طرح‌ها احتمال‌های شمول هدف با احتمال‌های از پیش تعیین‌شده برابرند. طرح نمونه‌گیری سیستماتیک که دارای حداقل تکیه‌گاه طرح است از کمترین میزان آنتروپی برخوردار است. جدول ۲ نشان می‌دهد، طرح‌هایی که حداکثر میزان آنتروپی دارند، از لحاظ اندازه‌ی فاصله‌ی هلینگر به یکدیگر بسیار نزدیک هستند. (جداول ۱ و ۲ در پیوست ارائه شده اند.)

مثال **۲.۴.** *جامعه‌ی معروف به سامفورد-هاجک با ۱۰ = N و اندازه‌ی نمونه‌ی ۵ = n و بردار احتمال‌های شمول هدف زیر را در نظر بگیرید: (۰/۲,۰/۲۵,۰/۳۵,۰/۴,۰/۵,۰/۵,۰/۵۵,۰/۶۵,۰/۷,۰/۹). آنتروپی طرح‌های مورد بررسی با بررسی تمام نمونه‌های ممکن و یا از طریق شبیه‌سازی محاسبه کرده و نتایج در جدول ۳ آورده شده است. همانند مثال قبلی مشاهده می‌شود، نمونه‌گیری پواسون شرطی تعدیل شده دارای حداکثر میزان آنتروپی است و چهار طرح نمونه‌گیری برتر نمونه‌گیری پواسون شرطی تعدیل‌شده، پارتو تعدیل‌شده، بریور و سامفورد است که طرح‌هایی با حداکثر میزان آنتروپی هستند و طرح نمونه‌گیری سیستماتیک کمترین میزان آنتروپی را دارد. جدول ۴ نشان می‌دهد، طرح‌هایی که حداکثر میزان آنتروپی دارند، دارای اندازه‌ی فاصله‌ی هلینگر بسیار نزدیک هستند. (جداول ۳ و ۴ در پیوست ارائه شده اند.)*

بحث و نتیجه‌گیری

برای انتخاب طرح نمونه‌گیری ملاک‌های بسیاری می‌تواند وجود داشته باشد. در این مقاله طرح‌های πps با معیار آنتروپی و فاصله هلینگر با یکدیگر مقایسه شدند و این نتیجه حاصل شد که چهار طرح نمونه‌گیری پواسون شرطی تعدیل‌شده، پارتو تعدیل‌شده، سامفورد و روش بریور طرح‌هایی با بیشترین میزان آنتروپی هستند. این طرح‌ها دقیق بوده و احتمال‌های شمول هدف با احتمال‌های از پیش تعیین شده برابرند. از لحاظ فاصله‌ی هلینگر نیز این طرح‌ها به یکدیگر بسیار نزدیک هستند. بنابراین می‌توان بر اساس کارایی و اجرای آسان و سریع از بین این طرح‌ها طرح مناسب را انتخاب نمود.

تقدیر و تشکر

نویسندگان از تمامی کسانی که در این تحقیق از راهنمایی آن‌ها بهره‌مند شده‌اند تشکر و قدردانی می‌کنند.

مراجع

[1] Aires, N. (1999). Algorithms to find exact inclusion probabilities for conditional Poisson sampling and Pareto πps sampling designs. Methodology and Computing in Applied Probability, 4, 457-469.

[2] Brewer, K. R. W. (1963). A model of systematic sampling with unequal probabilities. Australian Journal of Statistics, 5, 5-13.

[3] Brewer, K. R. W. (1975). A simple procedure for πpswor. Australian Journal of Statistics, 17, 166-172.

[4] Chen, S. X., Dempster, A. P. and Liu, J. S. (1994). Weighted finite population sampling to maximize entropy. Biometrika, 81, 457-469.

[5] Deville, J.-C. and Tille, Y. (1998). Unequal probability sampling without replacement through a splitting method. Biometrika, 85, 89-101.

[6] Hajek, J. (1964). Asymptotic theory of rejective sampling with varying probabilities from a finite population. Ann. Math. Statist., 35, 1491-1523.

[7] Hajek, J. (1981). Sampling from a finite population. Marcel Dekker, New York.

[8] Madow, W. G. (1949). On the theory of systematic sampling. Annals of Mathematical Statistics, 20, 333-354.

[9] Pea, J., Qualite, L. and Tille, Y. (2007). Systematic sampling is a minimum support design. Comput. Statist. Data Anal., 51, 5591-5602.

[10] Rosen, B. (1997). Asymptotic theory for order sampling. J. Statist. Plann. Inference, 62, 135-158.

[11] Rosen, B. (1997). On sampling with probability proportional to size. J. Statist. Plann. Inference, 62, 159-191.

[12] Sampford, M. R. (1967). On sampling without replacement with unequal probabilities of selection. Biometrika, 54, 499-513.

[13] Shannon, C. E. (1948). A mathematical theory of communication. Bell System Technical Journal, 27, 379-423 , 623-656.

[14] Tille, Y. (2006). Sampling algorithms. Springer series in statistics, Springer science, Business media, Inc., New York.

پیوست

جداول مربوط به این مقاله در ذیل ارائه شده است.

جدول ۱: آنتروپی طرح‌های نمونه‌گیری πps

طرح نمونه‌گیری	آنتروپی طرح
پواسون شرطی تعدیل‌شده	۲٫۷۱۵۱۷۶
پارتو تعدیل‌شده	۲٫۷۱۵۱۱۱
بریور	۲٫۷۱۵۰۷۸
سامفورد	۲٫۷۱۵۰۲۶
روش محوری	۲٫۷۰۹۲۰۰
پارتو	۲٫۶۹۵۵۵۸
پواسون شرطی	۲٫۵۸۱۵۵۲
خرد کردن به نمونه‌گیری تصادفی ساده	۱٫۰۹۸۶۱۲
سیستماتیک ترتیبی	۱٫۰۹۸۶۱۲

جدول ۲: فاصله‌ی هلینگر بین طرح‌های πps در جامعه‌ی TBM

طرح	ACP	$APar$	$Brewer$	$Sampf$	Piv	Par	$RSyst$	CP	$SSRS$	$Syst$
ACP	۰									
$APar$	۰/۰۰۴۰	۰								
$Brewer$	۰/۰۰۵۰	۰/۰۰۱۱	۰							
$Sampf$	۰/۰۰۶۱	۰/۰۰۲۱	۰/۰۰۱۴	۰						
Piv	۰/۰۳۶۲	۰/۰۴۰۰	۰/۰۴۰۵	۰/۰۴۱۹	۰					
Par	۰/۰۱۰۲	۰/۰۰۹۳	۰/۰۰۹۳	۰/۰۰۹۵	۰/۰۴۰۸	۰				
$RSyst$	۰/۰۹۵۴	۰/۰۹۹۰	۰/۰۹۹۴	۰/۱۰۰۸	۰/۰۵۹۸	۰/۰۹۸۷	۰			
CP	۰/۰۵۷۹	۰/۰۵۸۳	۰/۰۵۸۴	۰/۰۵۸۵	۰/۰۶۵۲	۰/۰۴۹۱	۰/۱۰۴۳	۰		
$SSRS$	۰/۲۱۸۴	۰/۲۲۲۳	۰/۲۲۳۲	۰/۲۲۴۴	۰/۱۸۶۵	۰/۲۲۲۵	۰/۱۴۷۴	۰/۲۲۱۳	۰	
$Syst$	۰/۷۴۴۶	۰/۷۴۳۲	۰/۷۴۲۹	۰/۷۴۲۵	۰/۷۵۴۵	۰/۷۴۳۲	۰/۷۶۹۲	۰/۷۴۶۶	۰/۸۲۶۹	۰

جدول ۳: آنتروپی طرح های πps.

طرح	آنتروپی طرح
پواسون شرطی تعدیل‌شده	۴/۷۲۶۹۹۰
پارتو تعدیل‌شده	۴/۷۲۶۹۶۳
بریور	۴/۷۲۶۹۵۵
سامفورد	۴/۷۲۶۹۵۰
محوری	۴/۷۲۵۵۰۰
پارتو	۴/۷۱۵۳۸۰
سیستماتیک تصادفی	۴/۷۱۱۵۱۹
پواسون شرطی	۴/۵۹۲۶۲۶
خرد کردن به نمونه‌گیری تصادفی ساده	۴/۴۹۰۰۸۶
سیستماتیک ترتیبی	۱/۹۰۰۲۲۶

جدول ۴: فاصله‌ی هلینگر بین طرح‌های πps در جامعه‌ی سامفورد-هاجک.

طرح	ACP	$APar$	$Brewer$	$Sampf$	Piv	$SSRS$	$Syst$
ACP	۰						
$Apar$	۰/۰۰۲۶	۰					
$Brewer$	۰/۰۰۳۰	۰/۰۰۴۴	۰				
$Sampf$	۰/۰۰۳۲	۰/۰۰۰۶	۰/۰۰۴۸	۰			
Piv	۰/۰۲۱۰	۰/۰۲۳۰	۰/۰۲۱۱	۰/۰۲۳۵	۰		
$SSRS$	۰/۲۳۹۶	۰/۲۴۱۹	۰/۲۳۹۶	۰/۲۴۲۴	۰/۲۲۴۰	۰	
$Syst$	۰/۸۵۳۴	۰/۸۵۳۳	۰/۸۵۳۳	۰/۸۵۳۲	۰/۸۵۲۸	۰/۸۶۶۷	۰