مجله‌ی بررسی‌های آمار رسمی ایران
سال۲۴، شماره‌ی ۲، پاییز و زمستان ۱۳۹۲، صص ۱۶۷-۱۸۴

براورد مجموع در آمارگیری‌های سه‌چارچوبی

طیبه چگینی* و حمیدرضا نواب‌پور

دانشگاه علامه طباطبایی

چکیده: در آمارگیری از جامعه‌های متناهی معمولاً یک نمونه‌ی احتمالاتی از یک چارچوب نمونه‌گیری که واحدهای جامعه‌ی هدف را پوشش می‌دهد انتخاب می‌شود. در برخی موارد، چارچوبی که در دسترس است به‌صورت کامل جامعه‌ی هدف را پوشش نمی‌دهد. لذا آماره‌های آمارگیری با استفاده از این چارچوب‌ها اغلب اریب هستند. از طرفی روزامدسازی چارچوبی که در دسترس است نیاز به صرف هزینه‌های زیاد دارد. از آمارگیری‌های چندچارچوبی با هدف افزایش پوشش جامعه و بهبود میزان اریبی و کاهش هزینه‌ها استفاده می‌شود. در آمارگیری‌های چندچارچوبی، نمونه‌ها به‌صورت مستقل از چندچارچوب که پوشش مناسبی از جامعه‌ی هدف ارایه می‌کنند و هم‌پوشانی دارند انتخاب می‌شوند. چندین براوردگر برای مجموع جامعه‌ای در آمارگیری‌های چندچارچوبی مطرح شده است. ما در این مقاله ضمن معرفی این براوردگرها، در یک مطالعه‌ی شبیه‌سازی کارایی براوردگرهای مختلف را با هم مقایسه می‌کنیم.

واژگان کلیدی: چارچوب نمونه‌گیری؛ آمارگیری‌های چندچارچوبی؛ براوردگر هارتلی؛ براوردگر فولر - بورمیستر؛ براوردگر ماکسیمم درستنمایی‌نما.

۱- مقدمه

در نمونه‌گیری به روش کلاسیک از جامعه‌ی متناهی، یک نمونه‌ی احتمالاتی از یک چارچوب نمونه‌گیری که شامل واحدهای جامعه‌ی هدف است، انتخاب می‌شود. ولی از آن‌جایی که معمولاً در عمل پوشش هیچ چارچوبی کامل نیست، راه مناسب‌تر آن است که نمونه‌های مستقلی از Q چارچوب نمونه‌گیری متفاوت، که اجتماع آن‌ها جامعه‌ی مورد نظر

---

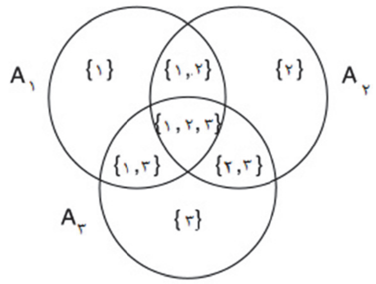
* نویسنده‌ی عهده‌دار مکاتبات

دریافت: ۱۳۹۲/۱۰/۸، پذیرش: ۱۳۹۳/۹/۳.

را پوشش می‌دهد، انتخاب شود و سپس اطلاعات به‌دست آمده از نمونه‌ها برای براورد کمیت‌های جامعه ترکیب شوند.
برای مثال یک آمارگیری در جامعه‌ی آمارشناسان ایالت‌های متحده را در نظر بگیرید. در این آمارگیری یکی از چارچوب‌های نمونه‌گیری می‌تواند فهرست عضوهای مستقیم انجمن امریکایی آمار (ASA) باشد که یک نمونه‌ی احتمالاتی از افرادی که عضو فهرست اعضای این انجمن هستند انتخاب می‌شود. بیش‌ترین افراد در این نمونه احتمالاً آمارشناس بوده و بنا بر این چارچوب مورد نظر مقرون به صرفه خواهد بود. با این حال بسیاری از آمارشناسان در ایالت‌های متحده عضو ASA نیستند، در نتیجه فهرست اعضای این انجمن، جامعه‌ی آمارشناسان ایالت‌های متحده را به‌طور کامل پوشش نمی‌دهد. پوشش این چارچوب را می‌توان با یک چارچوب دوم مانند فهرست عضوهای مستقیم مؤسسه‌ی آمار ریاضی (IMS) بهبود بخشید. به این ترتیب یک نمونه‌ی احتمالاتی از چارچوب A_1 (عضوهای مستقیم ASA) و مستقل از یک نمونه‌ی احتمالاتی از چارچوب A_2 (عضوهای مستقیم IMS) در نظر گرفت. بسیاری از آمارشناسان به هر دو سازمان تعلق دارند، بنا بر این این دو چارچوب هم‌پوشانی دارند. چارچوب A_1 و A_2 با هم، پوشش بهتری برای جامعه فراهم می‌کنند اما هنوز تمام آمارشناسان را شامل نمی‌شوند. در این حالت، یک آمارگیری سه‌چارچوبی می‌تواند مورد استفاده قرار گیرد. برای همین منظور چارچوب A_3 ممکن است یک چارچوب از تمام جامعه‌ی بزرگسالان باشد.

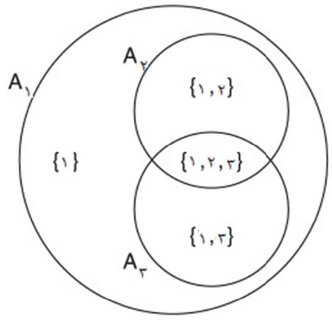
ممکن است این پرسش مطرح شود که چرا از همان ابتدا از چارچوب A_3 که شامل تمام جامعه است یک نمونه گرفته نمی‌شود؟ در واقع به دلیل هزینه‌ی زیاد آمارگیری از این چارچوب، از دو چارچوب ناقص دیگر استفاده شده است.

هاینس و پولاک استفاده از آمارگیری‌های چندچارچوبی را برای براورد اندازه‌ی جامعه مطرح کردند و به بررسی تناظر بین آمارگیری‌های چندچارچوبی و روش‌های گیر و بازگیر پرداختند [۳]. آن‌ها اطلاع‌های موجود از چارچوب‌های فهرستی ناقص و یک چارچوب ناحیه‌ای را برای براورد تعداد عقاب‌ها در یک ناحیه ترکیب کردند. ایچان و دنیس طرح چندچارچوبی را برای نمونه‌گیری از جامعه‌ی بی‌خانمان‌ها استفاده کردند که چارچوب‌های مورد استفاده در این طرح A_1 پناهگاه‌های بی‌خانمان‌ها، A_2 نوانخانه‌ها و A_3 مکان‌های خیابانی بودند [۷]. ساختار این طرح در شکل ۱ به تصویر کشیده شده است [۱۱].



شکل ۱- مثالی از یک آمارگیری سه‌چارچوبی

در جامعه‌های کمیاب نیز اغلب، استفاده از آمارگیری‌های چندچارچوبی می‌تواند نمونه‌های کاراتری به‌دست دهد. برای مثال، در مطالعه‌ی مشخصه‌ای از افراد مبتلا به آسم یکی از چارچوب‌ها می‌تواند چارچوبی باشد که در سرشماری سلامت از کل جامعه به‌دست آمده در حالی که می‌توان از دیگر چارچوب‌ها نظیر فهرست افراد در کلینیک‌های مربوط به یک بیماری خاص، مراکز درمانی و بیمارستان‌ها استفاده کرد. با آمارگیری‌های چندچارچوبی انتظار می‌رود که براوردهای دقیق‌تری از افراد مبتلا به بیماری آسم به‌دست آید، چرا که با سه چارچوب یا بیش‌تر، پوشش کامل‌تری از افراد مبتلا به بیماری آسم به‌دست خواهد آمد. ساختار این طرح در شکل ۲ نشان داده شده است. چارچوب $A_۱$ گرچه چارچوبی از کل واحدهای جامعه است اما ممکن است پوشش $A_۲$ و $A_۳$ در برخی بخش‌های جامعه کامل‌تر باشد.



شکل ۲- مثال دیگری از یک آمارگیری سه‌چارچوبی

اغلب یکی از هدف‌ها در تحلیل داده‌های آمارگیری‌های سه‌چارچوبی براورد مجموع جامعه‌ای Y است. در آمارگیری‌های چندچارچوبی، براوردگرهای مختلفی برای براورد

مجموع جامعه‌ای پیش‌نهاد شده است. هارتلی [۵، ۶] و فولر و بورمیستر [۲] به معرفی براوردگرهایی پرداختند که واریانس را در میان رده‌ای از براوردگرهای نااریب خطی Y مینیمم می‌کند. بانکی‌یر [۱] و کالتون و اندرسون [۸] براوردگرهای تک‌چارچوبی را در مشاهده‌هایی که بر اساس احتمال شمول‌شان در چارچوب‌های مختلف وزن‌دهی می‌شوند، گسترش دادند. اسکینر [۱۶] براوردگرهای نسبتی چنگک‌زنی را برای زمانی که از هر چارچوب نمونه‌های تصادفی انتخاب شوند، پیش‌نهاد کرد. اسکینر و رائو [۱۷] براوردگر ماکسیمم درست‌نمایی‌نما را برای Y در آمارگیری‌های دوچارجوبی با استفاده از طرح‌های پیچیده به‌دست آوردند. هر چند این براوردگر برای آمارگیری‌هایی با بیش از دو چارچوب نیز محاسبه می‌شود. لوهر و رائو [۹] کارایی براوردگرهای مجموع جامعه‌ای را در آمارگیری‌های دوچارجوبی مقایسه کردند. همچنین آن‌ها براوردگرهای خطی بهین و درست‌نمایی‌نما را برای چارچوب‌های چندگانه (با بیش از دو چارچوب) گسترش دادند.

در سال‌های اخیر براوردگرهای جدیدی نیز برای این آمارگیری‌ها پیش‌نهاد شده‌اند. از جمله مه‌کاتی [۱۴] براوردگر چندبارگی تک‌چارچوبی را برای این آمارگیری‌ها ارایه داد. لوهر و لو [۱۳] براوردگرهای جریان ناخالص در آمارگیری دوچارچوبی را مطرح ساختند. رائو و وو [۱۵] نیز روش تجربی درست‌نمایی‌نما را برای استنباط از آمارگیری‌های چندچارچوبی به کار بردند.

در بخش ۲ نمادها برای براورد مجموع جامعه‌ای در آمارگیری‌های چندچارچوبی معرفی می‌شوند. در بخش ۳ براوردگرها در آمارگیری‌های چندچارچوبی ارایه می‌شوند. در پایان با انجام یک مطالعه‌ی شبیه‌سازی، به مقایسه‌ی این براوردگرها در یک آمارگیری سه‌چارچوبی پرداخته می‌شود.

۲- نمادها

در آمارگیری‌های چندچارچوبی، نمونه‌های احتمالاتی به‌صورت مستقل از چارچوب‌های $A_1, \dots, A_Q$، $Q \geq 2$ انتخاب می‌شوند. فرض کنید $\mathrm{F} = \left\{1, \dots, Q\right\}$ مجموعه‌ی اندیس چارچوب‌ها باشد، به‌طوری که اجتماع این چارچوب‌ها جامعه‌ی مورد نظر را بپوشاند. با

Q چارچوب $2^Q - 1$ حوزه‌ی مجزای ممکن وجود دارند که توسط زیرمجموعه‌های F تعریف می‌شوند. برای $K \subseteq F$ حوزه‌ی تعریف‌شده توسط K چنین است:

$$D_K = \left(\bigcap_{i \in K} A_i\right) \bigcap \left(\bigcap_{j \notin K} A_j^c\right)$$

که A_j^c مکمل مجموعه‌ی A_j است. تعداد چارچوب‌های سهیم در ساختن حوزه‌ی K با $|K|$ نشان داده می‌شود. فرض کنید $N^{(q)}$ تعداد واحدهای جامعه در چارچوب A_q و N_K تعداد واحدهای جامعه در حوزه‌ی تعریف شده توسط K باشد. از آن‌جایی که حوزه‌ها متداخل نیستند اندازه‌ی جامعه از رابطه‌ی $\sum_{K \subseteq F} N_K = N$ محاسبه می‌شود.

دو مثال از آمارگیری سه‌چارچوبی در شکل‌های ۱ و ۲ نشان داده شده‌اند. شکل ۱، یک طرح سه‌چارجوبی را نشان می‌دهد که در آن هر یک از چارچوب‌ها ناقص هستند. حوزه‌های $\{1\}$، $\{2\}$ و $\{3\}$ به‌ترتیب تنها در چارچوب‌های A_1، A_2 و A_3، حوزه‌های $\{1,2\}$، $\{1,3\}$ و $\{2,3\}$ دقیقاً در دو چارچوب ظاهر شده‌اند و حوزه‌ی $\{1,2,3\}$ اشتراک سه چارچوب است. شکل ۲ نیز یک طرح سه‌چارجوبی است با این تفاوت که چهار حوزه دارد. چارچوب A_1 کامل است اما چارچوب‌های A_2 و A_3 ناقص هستند و با هم تداخل دارند.

مجموع جامعه‌ای Y به‌صورت جمع مجموع جامعه‌ای حوزه‌های مجزا نوشته می‌شود:

$$Y = \sum_{K \subseteq F} Y_K$$

که در آن $Y_K = \sum_{i=1}^{N} \delta_i(K) y_i$. اگر واحد مشاهده‌ای i عضو حوزه‌ی تعریف‌شده توسط K باشد $\delta_i(K)$ برابر با یک خواهد بود و در غیر این صورت مقدار صفر می‌گیرد. در حالت خاص، که تعداد واحدهای جامعه در حوزه‌ی D_K به‌ازای همه‌ی iها برابر یک باشد، با در نظر گرفتن $y_i = 1$ اندازه‌ی حوزه از رابطه‌ی $N_K = \sum_{i=1}^{N} \delta_i(K)$ به‌دست می‌آید. مشکلی که در براورد مجموع جامعه‌ای Y وجود دارد این است که در هر حوزه مؤلفه‌های Y_K کم براورد می‌شوند، به‌همین دلیل از براورد $|K|$ در چارچوب‌های نمونه‌گیری مختلف استفاده می‌شود.

فرض کنید S_q نشان‌دهنده‌ی نمونه‌ای احتمالاتی از چارچوب A_q برای $q = ۱, \ldots, Q$ باشد. در آمارگیری‌های چندچارچوبی چون نمونه‌ها از Q چارچوب به‌طور مستقل انتخاب می‌شوند، به تعداد $|K|$ براوردگر مستقل از Y_Kها وجود دارد. برای $q \in K$، براوردگر مجموع با استفاده از نمونه‌های چارچوب A_q به‌صورت زیر براورد می‌شود:

$$\hat{Y}_K^{(q)} = \sum_{i \in S_q} w_i^{(q)} \delta_i(K) y_i \qquad (۱)$$

وزن‌های $w_i^{(q)}$ می‌توانند وارون احتمال شمول یا وزن هایک [۴] باشد. به‌طور مشابه براوردگر اندازه‌ی حوزه‌ای این چنین تعریف می‌شود:

$$\hat{N}_K^{(q)} = \sum_{i \in S_q} w_i^{(q)} \delta_i(K). \qquad (۲)$$

فرض کنید در تمام حالت‌ها واحدها را در هر یک از آمارگیری‌ها می‌توان تشخیص داد. البته این یک فرض قوی است که ممکن است همیشه درست نباشد [۱۰].

۳- براوردگرهای مجموع جامعه‌ای در آمارگیری‌های چندچارچوبی

در این بخش نحوه‌ی محاسبه‌ی براوردگرهای مجموع جامعه‌ای در آمارگیری‌های چندچارچوبی ارایه شده است. این براوردگرها عبارت‌اند از: براوردگرهای خطی بهین، براوردگر ماکسیمم درست‌نمایی‌نما، براوردگر تک‌چارچوبی و براوردگر تعدیل وزنی.

۱-۳- براوردگرهای خطی بهین

در این بخش ابتدا براوردگرهای بهین (هارتلی و فولر-بورمیستر) برای طرح‌های آمارگیری عمومی معرفی می‌شوند. حوزه‌های ناخالی با $K_۱, \ldots, K_d$ برچسب‌گذاری می‌شوند که البته ممکن است این حوزه‌های ناخالی از ساختار چارچوب‌ها مانند شکل ۱ شناسایی شوند یا زمانی که اندازه‌ی نمونه به حد کافی بزرگ باشد احتمال گزینش یک اندازه‌ی نمونه‌ی صفر از حوزه‌های ناخالی ناچیز باشد. فرض کنید $\boldsymbol{\theta}_q = \left(\boldsymbol{\theta}_{K_۱}^{(q)}, \ldots, \boldsymbol{\theta}_{K_d}^{(q)}\right)^T$ و $\boldsymbol{\theta} = \left(\theta_۱^T, \ldots, \theta_Q^T\right)^T$ بردارهایی از ثابت‌ها و مشخص باشند. همچنین براورد مجموع جامعه‌ای و اندازه‌ی جامعه‌ای در qامین چارچوب به‌صورت $\hat{\boldsymbol{Y}}_{\boldsymbol{q}} = \left(\hat{Y}_{K_۱}^{(q)}, \ldots, \hat{Y}_{K_d}^{(q)}\right)^T$ و

$\widehat{\boldsymbol{N}}_{\boldsymbol{q}} = \left(\widehat{N}_{K_1}^{(q)}, \dots, \widehat{N}_{K_d}^{(q)}\right)^T$ تعریـف می‌شـوند کـه اگـر حـوزه‌ی K_i بخشـی از چـارچوب A_q نباشـد یعنـی $q \notin K_i$ آن‌گـاه $\widehat{N}_{K_i}^{(q)} = \hat{Y}_{K_i}^{(q)} = ۰$ و $\theta_{K_i}^{(q)} = ۰$. مـاتریس قطـری $d \times d$، $\boldsymbol{\Gamma}_q$ طوری تعریف می‌شود که iامین عنصر قطری‌اش یک است اگر حـوزه‌ی K_i بخشـی از چارچوب A_q باشد $(q \in K_i)$ و در غیر این‌صورت صفر است. نماد ۱_d بردار dبُعـدی از یک‌ها در نظر گرفته می‌شود. بنا بر این براورد هارتلی Y برای مقدارهایی از θ برابر است با:

$$\widehat{\boldsymbol{Y}}_\theta = \sum_{q=۱}^{Q} \boldsymbol{\theta}_q^T \boldsymbol{\Gamma}_q \widehat{\boldsymbol{Y}}_q \qquad (۳)$$

اگر بردار ثابت $\boldsymbol{\theta}$ در قید زیر صدق کند

$$\sum_{q=۱}^{Q} \boldsymbol{\Gamma}_q \boldsymbol{\theta}_q = ۱_d \qquad (۴)$$

و $\mathrm{E}\left[\hat{Y}_K^{(q)}\right] = Y_K$ بـرای هـر حـوزه‌ی K و چـارچوب A_q برقـرار باشـد، آن‌گـاه $\hat{Y}_\theta$ یـک براوردگر نااریب برای مجموع جامعه‌ای Y خواهد بود.

براوردگر هارتلی تعمیم‌یافته از جایگزینی θ_H به‌جای θ حاصل می‌شود به‌طوری که θ_H واریانس $\hat{Y}_\theta$ را تحت قید (۴) مینیمم می‌کند. با تعریف $\boldsymbol{C}_q$ به‌عنوان ماتریس کوواریـانس $\widehat{\boldsymbol{Y}}_q$ و مینیمم کردن واریانس $\hat{Y}_\theta$ از رابطه‌ی:

$$\mathrm{V}\left(\widehat{\boldsymbol{Y}}_\theta\right) = \sum_{q=۱}^{Q} \boldsymbol{\theta}_q^T \boldsymbol{\Gamma}_q \boldsymbol{C}_q \boldsymbol{\Gamma}_q \boldsymbol{\theta}_q$$

می‌توان θ_{H} را از حل معادله‌ی زیر به‌دست آورد:

$$\mathbf{A\theta} = \begin{bmatrix} ۰_{\mathrm{dQ}}^{\mathrm{T}} & ۱_{\mathrm{d}}^{\mathrm{T}} \end{bmatrix}^{\mathrm{T}} \qquad (۵)$$

کـه در آن ۰_{dQ} یـک بـردار dQبُعـدی سـتونی از صفرهاسـت، $\mathbf{A}$ یـک مـاتریس بلـوکی $d(Q+۱) \times dQ$بُعدی است به‌طوری که (i,j)امین بلوک آن برابر است با:

$$A_{i,j}=\begin{cases}\Gamma_i\left(\sum_{q=1}^{Q}\Gamma_q\right)^{-1}\left(\boldsymbol{I}-\sum_{q=1}^{Q}\Gamma_q\right)E_i & i=j\\ \Gamma_i\left(\sum_{q=1}^{Q}\Gamma_q\right)^{-1}E_j & i\neq j\,, i\leq Q\\ \Gamma_j & i=Q+1\end{cases}$$

همچنین $\boldsymbol{I}$ ماتریس همانی $d\times d$ و $\boldsymbol{E}_i=\boldsymbol{\Gamma}_i\boldsymbol{C}_i\boldsymbol{\Gamma}_i$ است.

حال به سراغ براوردگر بهین دیگر یعنی براوردگر فولر-بورمیستر می‌رویم. برای تعمیم براوردگر فولر-بورمیستر، از بردارهای اندازه‌ی حوزه‌های براوردشده‌ی $\widehat{\boldsymbol{N}}^{(q)}$ استفاده می‌شود. فرض کنید $\boldsymbol{\beta}=\left(\boldsymbol{\beta}_1^T,\dots,\boldsymbol{\beta}_Q^T\right)^T$ یک بردار $2dQ\times 1$ از پارامترها باشد و ماتریس $\mathbf{G}_q$ به‌صورت $\boldsymbol{G}_q=\operatorname{diag}(\boldsymbol{\Gamma}_q,\boldsymbol{\Gamma}_q)$ تعریف شود. آن‌گاه براوردگر فولر-بورمیستر برابر خواهد بود با:

$$\hat{Y}_{FB}(\boldsymbol{\beta})=\sum_{q=1}^{Q}\boldsymbol{\beta}_q^T\boldsymbol{G}_q\left[\widehat{\boldsymbol{Y}}_q^T\quad \widehat{\boldsymbol{N}}_q^T\right]^T \qquad (۶)$$

که $\boldsymbol{\beta}$ در قید زیر صدق می‌کند:

$$\sum_{q=1}^{Q}\boldsymbol{G}_q\boldsymbol{\beta}_q=\left[1_d^T\quad 0_d^T\right]^T \qquad (۷)$$

با تعریف ماتریس کوواریانس

$$\mathbf{D}_q=\operatorname{cov}\begin{bmatrix}\widehat{\mathbf{Y}}_q\\ \widehat{\mathbf{N}}_q\end{bmatrix}=\begin{bmatrix}\boldsymbol{C}_q & \boldsymbol{D}_{q12}\\ \boldsymbol{D}_{q12}^T & \boldsymbol{D}_{q22}\end{bmatrix}$$

واریانس براوردگر فولر-بورمیستر به‌صورت زیر در می‌آید:

$$\mathrm{V}\left(\hat{Y}_{FB}\right)=\sum_{q=1}^{Q}\boldsymbol{\beta}_q^T\boldsymbol{G}_q\boldsymbol{D}_q\boldsymbol{G}_q\boldsymbol{\beta}_q$$

مشابه (۵) واریانس براوردگر فولر-بورمیستر نسبت به قید (۷) مینیمم می‌شود زمانی که $\boldsymbol{\beta}$ در معادله‌ی زیر صدق کند:

$$\boldsymbol{B}\boldsymbol{\beta}=\left[0_{2dQ}^T\quad 1_d^T\quad 0_d^T\right]^T$$

کـه در آن بلوک‌هـای $۲dQ \times ۲d(Q+۱)$ بُعـدی مـاتریس $\boldsymbol{B}$ مشـابه مـاتریس $\mathbf{A}$ در برآوردگر هارتلی تعریف می‌شوند، با این تفاوت که $\boldsymbol{G}_i$ جایگزین $\boldsymbol{\Gamma}_i$ خواهد شد [۱۰].

۲-۳- برآورد ماکسیمم درست‌نمایی‌نما در نمونه‌گیری‌های تصادفی ساده

یک نمونه‌ی تصادفی ساده به انـدازه‌ی $n^{(q)}$ از چـارچوب A_q، بـرای $q = ۱, \dots, \mathrm{Q}$ را در نظر بگیرید. فرض کنید $\boldsymbol{f} = \left(f^{(۱)}, \dots, f^{(Q)}\right)$ بردار Qبُعدی از کسرهای نمونه‌گیری بـا $f^{(q)} = \frac{n^{(q)}}{N^{(q)}}$ باشد. لازم به ذکر است که برای نمونـه‌گیری تصـادفی سـاده وزن نمونـه‌گیری واحـد iام جامعـه $\left(i = ۱, \dots, N^{(q)}\right)$ عکـس کسـر نمونـه‌گیری اسـت. بـه‌عبـارت دیگـر $w_i^{(q)} = \frac{۱}{f^{(q)}}$ بنـا بـر ایـن $n_K^{(q)} = \widehat{N}_K^{(q)} f^{(q)}$ انـدازه‌ی نمونـه‌ای از چـارچوب A_q در حوزه‌ی K است که $\widehat{N}_K^{(q)}$ در (۲) تعریف شده است. در آمـارگیری چنـدچارچوبی می‌تـوان مجموع جامعه‌ای را به‌صورت زیر نوشت:

$$Y = \sum_{K \subseteq F} Y_K = \boldsymbol{N}^T \overline{\boldsymbol{Y}}$$

کــه در آن $\boldsymbol{N} = \left(N_{K_۱}, \dots, N_{K_d}\right)^T$، $\boldsymbol{Y} = \left(Y_{K_۱}, \dots, Y_{K_d}\right)^T$ و $\overline{\boldsymbol{Y}}$ بــردار متنــاظر بــا میانگین حوزه‌ها هستند. لوهر و رائو [۱۰] برآوردهای ماکسیمم درست‌نمایی را بـرای $\overline{\boldsymbol{Y}}$ و $\boldsymbol{N}$ محاسبه کردند به‌طوری که برآورد ماکسیمم درست‌نمایی‌نما برای $\boldsymbol{Y}$ برابر است با:

$$\hat{Y} = \widehat{\boldsymbol{N}}^T \widehat{\overline{\boldsymbol{Y}}} = \sum_{K \subseteq F} \widehat{N}_K \widehat{\overline{Y}}_K \qquad (۸)$$

برآوردگر $\bar{y}_K^{(q)} = \sum_{i \in S_q} \delta_i(K) y_i / n_K^{(q)}$، تحت شرایط نظم و با استفاده از قضیه‌ی حد مرکزی، برای نمونه‌های به اندازه‌ی کافی بزرگ، دارای توزیع تقریبی نرمال با میـانگین $\overline{Y}_K$ و واریانس $\sigma_K^۲ / n_K^{(q)}$ اسـت. همچنـین بـرای هـر q، تحـت نمونـه‌گیری تصـادفی سـاده بـا کسرهای نمونه‌گیری کوچک، $\left\{n_K^{(q)}\right\}$ به‌طور تقریبی دارای توزیـع چندجملـه‌ای بـا انـدازه‌ی $n^{(q)}$ و احتمال موفقیت $N_K / N^{(q)}$ برای حوزه‌ی K در چارچوب A_q اسـت. بنـا بـر ایـن تابع درست‌نمایی $(\overline{\boldsymbol{Y}}, \boldsymbol{N})$ با $L(\overline{\boldsymbol{Y}}, \boldsymbol{N}) = L_۱(\overline{\boldsymbol{Y}}) L_۲(\boldsymbol{N})$ تقریب زده می‌شود که در آن c مقدار ثابت است.

$$\log L_{۱}(\bar{Y}) = c - \frac{۱}{۲}\log \sigma_K^{۲} - \sum_{q=۱}^{Q} \sum_{K:q\in K} \frac{n_K^{(q)}\left(\bar{y}_K^{(q)} - \bar{Y}_K\right)^{۲}}{۲\sigma_K^{۲}} \tag{۹}$$

و

$$\begin{aligned} \log L_{۲}(N) &= c + \sum_{q=۱}^{Q} \sum_{K:q\in K} n_K^{(q)} \log N_K \\ &= c + \sum_{q=۱}^{Q} \sum_{K:q\in K} \widehat{N}_K^{(q)} f^{(q)} \log N_K \end{aligned} \tag{۱۰}$$

با در نظر گرفتن قید

$$N^{(q)} = \sum_{K:q\in K} N_K, \qquad q = ۱, \dots, Q \tag{۱۱}$$

و با ماکسیمم کردن رابطه‌ی (۹) براورد $\bar{Y}_k$ به‌صورت زیر به‌دست می‌آید:

$$\hat{\bar{Y}}_K = \frac{\sum_{q\in K} n_K^{(q)} \bar{y}_K^{(q)}}{\sum_{q\in K} n_K^{(q)}} = \frac{\sum_{q\in K} f^{(q)} \hat{Y}_K^{(q)}}{\sum_{q\in K} f^{(q)} \widehat{N}_K^{(q)}}. \tag{۱۲}$$

برای ماکسیمم کردن رابطه‌ی (۱۰)، M را یک ماتریس $d \times Q$ در نظر بگیرید که (i,j)امین درایه‌ی آن یک است اگر $j \in K_i$ باشد و سایر جاها صفر است. همچنین فرض کنید که $\widehat{\boldsymbol{H}}$ یک $d \times Q$ به‌صورت زیر باشد، در این صورت:

$$\widehat{\boldsymbol{H}}_{i,j} = \begin{cases} \widehat{N}_{K_i}^{(q)} & q \in K_i \quad \text{اگر} \\ ۰ & \text{در غیر این‌صورت} \end{cases}$$

که برای نمونه‌گیری تصادفی ساده، $\widehat{N}_K^{(q)} = \left(N^{(q)}/n^{(q)}\right) n_K^{(q)}$ است. قید (۱۱) به این معنی است که برداری که رابطه‌ی (۱۰) را ماکسیمم می‌کند یعنی $\widehat{\boldsymbol{N}}$ باید در $\boldsymbol{M}^{\boldsymbol{T}}\widehat{\boldsymbol{N}} = \boldsymbol{h}$ صدق کند که در این رابطه $\widehat{\boldsymbol{N}}$ حل معادله‌ی زیر است:

$$\begin{bmatrix} (\boldsymbol{I} - \boldsymbol{M}\boldsymbol{M}^{+})(\text{diag}\boldsymbol{N})^{-۱}\widehat{\boldsymbol{H}}\boldsymbol{f} \\ \boldsymbol{M}^{\boldsymbol{T}}\widehat{\boldsymbol{N}} - \boldsymbol{h} \end{bmatrix} = ۰_{Q+\text{d}} \tag{۱۳}$$

در معادله‌ی بالا $\boldsymbol{h} = \left(N^{(۱)}, \dots, N^{(Q)}\right)^{T}$ و $\boldsymbol{M}^{+}$ وارون تعمیم‌یافته‌ی مور-پن‌رز ماتریس $\boldsymbol{M}$ است. اگر $\boldsymbol{M}$ وارون‌پذیر باشد آن‌گاه $\widehat{\boldsymbol{N}} = \boldsymbol{M}^{-\boldsymbol{T}}\boldsymbol{h}$. نظام معادله‌های (۱۳) نسبت به

$\widehat{\boldsymbol{N}}$ خطی نیست. لوهر و رائو [۱۰] با استفاده از قضیه‌ی نگاشت ضمنی یک تقریب خطی $\widetilde{\boldsymbol{N}}$ برای $\widehat{\boldsymbol{N}}$ یافتند. براوردگر خطی $\widetilde{\boldsymbol{N}}$ حل معادله‌ی زیر است:

$$\begin{bmatrix}(\boldsymbol{I}-\boldsymbol{M}\boldsymbol{M}^{+})(\mathrm{diag}\boldsymbol{N})^{-1}(\mathrm{diag}\boldsymbol{M}\boldsymbol{f})\\ \boldsymbol{M}^{\boldsymbol{T}}\end{bmatrix}\widetilde{\boldsymbol{N}}=\begin{bmatrix}(\boldsymbol{I}-\boldsymbol{M}\boldsymbol{M}^{+})(\mathrm{diag}\boldsymbol{N})^{-1}\widehat{\boldsymbol{H}}\boldsymbol{f}\\ \boldsymbol{h}\end{bmatrix} \tag{۱۴}$$

معادله‌ی (۱۴) برای یافتن واریانس مجانبی $\widehat{\boldsymbol{N}}$ که با واریانس مجانبی $\widetilde{\boldsymbol{N}}$ یکسان است مورد استفاده قرار می‌گیرد. همچنین این معادله برای محاسبه‌ی $\widehat{\boldsymbol{N}}$ به‌کمک روش‌های تکراری به‌کار می‌رود. به این صورت که با در نظر گرفتن یک بردار از $\widehat{\boldsymbol{N}}$، براوردهای آغازین برای $\boldsymbol{N}$ و $\widetilde{\boldsymbol{N}}_{K+1}$ حل معادله‌ی زیر باشد:

$$\begin{bmatrix}(\boldsymbol{I}-\boldsymbol{M}\boldsymbol{M}^{+})\left(\mathrm{diag}\widetilde{\boldsymbol{N}}_{K}\right)^{-1}(\mathrm{diag}\boldsymbol{M}\boldsymbol{f})\\ \boldsymbol{M}^{\boldsymbol{T}}\end{bmatrix}\widetilde{\boldsymbol{N}}_{K+1}=\begin{bmatrix}(\boldsymbol{I}-\boldsymbol{M}\boldsymbol{M}^{+})\left(\mathrm{diag}\widetilde{\boldsymbol{N}}_{K}\right)^{-1}\widehat{\boldsymbol{H}}\boldsymbol{f}\\ \boldsymbol{h}\end{bmatrix}$$

به این ترتیب با استفاده از این معادله می‌توان اندازه‌ی جامعه‌ای، $\boldsymbol{N}$ را براورد کرد. اگر براورد آغازین $\widehat{\boldsymbol{N}}$ قابل قبول باشد، آن‌گاه براورد حاصل از این روش به‌سرعت به $\widehat{\boldsymbol{N}}$ همگرا خواهد شد. در عمل مقدار آغازین را از متوسط $\widehat{N}_{K}^{(q)}$ها یا از چارچوب با بزرگ‌ترین اندازه‌ی نمونه در حوزه‌ی K محاسبه می‌کنند [۱۰].

۳-۳- روش‌های براورد تک‌چارچوبی

روش‌های تک‌چارچوبی توسط بانکی‌یر [۱] و کالتون و اندرسون [۸] در موقعیت‌های دوچارچوبی پیش‌نهاد شدند. در این روش مشاهده‌ها داخل یک چارچوب تکی ترکیب و وزن‌ها در حوزه‌ی مشترک تعدیل می‌شوند. در این بخش روش تک‌چارچوبی به‌منظور براورد مجموع جامعه‌ای با بیش از دو چارچوب به‌کار برده می‌شود.

در براوردگر بانکی‌یر [۱] لازم است که واحدها از نمونه‌های مختلف متمایز باشند. برای گسترش این براوردگر به چارچوب‌های چندگانه، قرار دهید:

$$\widetilde{w}_{i,S}^{(q)}=\sum_{K\subseteq F}\delta_{i}(K)\left(\sum_{j\in K}\frac{1}{w_{i}^{(j)}}\right)^{-1} \tag{۱۵}$$

که $w_{i}^{(j)}$ وزن واحد مشاهده‌ی iام از چارچوب j است. با وزن تعدیل‌شده‌ی بالا براوردگر تک‌چارچوبی برابر است با:

$$\hat{Y}_S = \sum_{q=1}^{Q} \sum_{i\in S_q} \widetilde{w}_{i,S}^{(q)} y_i \qquad (۱۶)$$

اگر هر نمونه خودوزن باشد (یعنـی $w_i^{(q)} = w_j^{(q)} = w^{(q)}$ بـرای همـه‌ی i، j و qهـا)، آن‌گاه $\hat{Y}_S$ می‌تواند به شکل (۳) نوشته شود. در این حالـت بـا اسـتفاده از فرمول‌هـای (۱)، (۱۵) و (۱۶) براوردگـر تک‌چـارچوبی در آمارگیری‌هـای چنـدچارچوبی به‌صـورت زیـر محاسبه می‌شود:

$$\hat{Y}_S = \sum_{q=1}^{Q} \sum_{K:q\in K} \left(w^{(q)} \sum_{j\in K} \frac{1}{w^{(j)}} \right)^{-1} \sum_{i\in S_q} w^{(q)} \delta_i(K) y_i = \sum_{q=1}^{Q} \theta_{q,S}^T \Gamma_q \hat{Y}_q$$

با

$$\boldsymbol{\theta}_{q,S}^T = \left[\left(w^{(q)} \sum_{j\in K_1} \frac{1}{w^{(j)}} \right)^{-1}, \dots, \left(W^{(q)} \sum_{j\in K_d} \frac{1}{w^{(j)}} \right)^{-1} \right]$$

که $\boldsymbol{\theta}_{q,S}^T$ پارامتر براوردگر تک‌چارچوبی یا همان وزن‌های تعدیل‌یافته بـرای چـارچوب qام است. هنگامی که اندازه‌ی چارچوب‌هـا معلـوم باشـند براوردگـر تک‌چـارچوبی می‌توانـد بـا استفاده از براورد نسبتی چنگک‌زنی [۱ و ۱۷]، یا براورد رگرسیونی تعدیل شود. زمانی که همه‌ی آمارگیری‌هـای مجـزا خـودوزن باشـند، براوردگـر تک‌چـارچوبی بـرای هـر تعـدادی از چارچوب‌ها نیز محاسبه خواهد شد.

۴-۳- براوردگر تعدیل وزنی با سه یا بیش از سه چارچوب

در حالـت کلـی فـرض کنیـد Q چـارچوب $A_1, \dots, A_Q$ وجـود دارنـد. نمونـه‌های احتمـالاتی مسـتقل از چـارچوب A_q بـا S_q بـرای $q = 1, \dots, Q$ نشـان داده می‌شـود. واحـد iام در نمونـه‌ی S_q دارای احتمـال شـمول $\pi_i^{A_q}$ و وزن نمونـه‌گیری $w_i^{A_q}$ اسـت. در مجمـوع D حوزه‌ی مجزا وجود دارد.

براوردگر مجموع جامعه‌ای با استفاده از وزن‌های تعدیل‌یافته به شکل زیر است:

$$\hat{Y} = \sum_{i\in S_q}^{Q} m_i^{A_q} w_i^{A_q} y_i$$

که در آن m_i^{Aq} عامل تعدیل وزن برای مشاهده‌ی iام در S_q است. اگر حوزه‌ی d بخشی از چارچوب A_q نباشد، $m^{(A_q,d)}$ (عامل تعدیل وزن برای هر چارچوب و حوزه) برابر با صفر است. البته تعدیل وزن مورد نظر باید در دو قید $m^{(A_q,d)} \geq ۰$ و $\sum_{q=۱}^{Q} m^{(A_q,d)} = ۱$ برای $d = ۱, \dots, D$ صدق کند. بنا بر این هنگامی که مشاهده‌ی iام از S_q در حوزه‌ی d است $m_i^{A_q} = m_i^{(A_q,d)}$ [۱۲].

برای مثال این براوردگر در آمارگیری‌های سه‌چارچوبی با استفاده از میانگین براوردهای مجموع جامعه‌ای حوزه‌ها به‌صورت زیر محاسبه می‌شود:

$$\hat{Y}_{ave} = \hat{Y}_a^A + \hat{Y}_b^B + \hat{Y}_c^C + \frac{۱}{۲}\left(\hat{Y}_{ab}^A + \hat{Y}_{ab}^B\right) + \frac{۱}{۲}\left(\hat{Y}_{ac}^A + \hat{Y}_{ac}^C\right) + \frac{۱}{۲}\left(\hat{Y}_{bc}^B + \hat{Y}_{bc}^C\right) + \frac{۱}{۳}\left(\hat{Y}_{abc}^A + \hat{Y}_{abc}^B + \hat{Y}_{abc}^C\right)$$

این براوردگر ساده‌ترین روش برای یافتن براورد مجموع جامعه‌ای در آمارگیری‌های سه‌چارچوبی است که روش تعدیل وزنی نام دارد.

۴- مطالعه‌ی شبیه‌سازی

در این بخش با استفاده از نرم‌افزار R، به روش شبیه‌سازی مونته کارلویی براوردگرهای هارتلی، فولر-بورمیستر، ماکسیمم درست‌نمایی‌نما، تک‌چارچوبی و تعدیل وزنی برای مجموع جامعه‌ای در آمارگیری سه‌چارچوبی مقایسه می‌شوند.
مشاهده‌های جامعه‌ای متناهی از آمارگیری سه‌چارچوبی برای حوزه‌ی Kام با استفاده از مدل احتمالاتی تصادفی زیر تولید شده است:

$$y_i = \mu_K + \varepsilon_i, \quad i = ۱, \dots, N_K\,, K = K_۱, \dots, K_d$$

که ε_i دارای توزیع نرمال با میانگین صفر و واریانس یک است. میانگین حوزه‌ی Kام، μ_K و اندازه‌ی جامعه‌ای هر حوزه با N_K نشان داده شده است. با اندازه‌ی جامعه‌ای معلوم

چارچوب‌ها و میانگین‌های مشخص، از هر سه چارچوب نمونه‌های تصادفی ساده و مستقل از هم گزینش شده‌اند.

برای انجام شبیه‌سازی، نمونه‌گیری را ۱۰۰۰ بار تکرار کرده و از هر نمونه ریشه‌ی دوم میانگین توان‌های دوم خطای تجربی برای براوردگرهای مجموع جامعه‌ای محاسبه شده است. همچنین به‌منظور ارزیابی اثر اندازه‌ی نمونه‌ای بر میانگین توان دوم خطای براوردگرها برای هر سه چارچوب اندازه‌های نمونه ۱۰۰، ۲۰۰ و ۵۰۰ در نظر گرفته شده است. نتیجه‌های شبیه‌سازی به ازای اندازه‌ی جامعه‌ای حوزه‌های مختلف و میانگین‌های متفاوت ارایه شده است.

یکی از معیارها برای بیان میزان درستی یک براوردگر، میانگین توان‌های دوم خطای تجربی است. با استفاده از شبیه‌سازی مونته کارلو، میانگین توان‌های دوم خطای تجربی برای تکرارهای مختلف نمونه‌گیری برابر است با:

$$\mathrm{EMSE} = \frac{۱}{B}\sum_{i=۱}^{B}(\hat{t}_i - t)^۲$$

که $\hat{t}_i$ براورد مجموع جامعه‌ای برای هر نمونه و t مجموع جامعه و معلوم است. ریشه‌ی دوم میانگین توان دوم خطای تجربی ($\sqrt{\mathrm{EMSE}}$) براوردگرهای مجموع جامعه‌ای، در دو جامعه با اندازه‌ی جامعه‌ای حوزه‌ها و میانگین‌های متفاوت محاسبه و در زیر ارایه شده است. با توجه به شکل ۱، ترتیب حوزه‌ها در هر دو جامعه از چپ به راست به‌صورت $\{۱\}$، $\{۱,۲\}$، $\{۲\}$، $\{۱,۲,۳\}$، $\{۱,۳\}$، $\{۲,۳\}$ و $\{۳\}$ است که اندازه‌ی جامعه‌ای در حوزه‌های $\{۱\}$، $\{۱,۲\}$ و $\{۱,۳\}$ در جامعه‌ی دوم افزایش داده شده است.

جامعه‌ی اول:

$$\boldsymbol{N^T} = (۲۰۰۰,۲۰۰۰,۲۰۰۰,۲۰۰۰,۲۰۰۰,۲۰۰۰,۲۰۰۰),\ \boldsymbol{\mu}^\mathrm{T} = (۱۱,۱۴,۱۲,۱۷,۱۶,۱۸,۱۹)$$

جامعه‌ی دوم:

$$\boldsymbol{N^T} = (۶۰۰۰,۶۰۰۰,۲۰۰۰,۲۰۰۰,۶۰۰۰,۲۰۰۰,۲۰۰۰),\ \ \boldsymbol{\mu}^\mathrm{T} = (۱۵,۱۰,۲۰,۱۷,۱۲,۱۸,۱۴)$$

با توجه به جدول ۱، براوردگر ماکسیمم درست‌نمایی‌نما کم‌ترین $\sqrt{\mathrm{EMSE}}$ را در میان سایر روش‌ها دارد. براوردگر تک‌چارچوبی و روش تعدیل وزنی تقریباً با هم برابرند. با توجه به این‌که نمونه‌های تصادفی از هر یک از جامعه‌ی چارچوب‌ها انتخاب شده‌اند در نتیجه وزن‌های نمونه‌گیری برای واحدهای جامعه‌ی چارچوب‌ها درون هر حوزه یکسان است.

جدول ۱- ریشه‌ی دوم میانگین توان دوم خطای تجربی براوردگرهای ۵گانه‌ی جامعه‌ی اول

$n^{(۱)}$	$n^{(۲)}$	$n^{(۳)}$	WA*	Hart	FB	PML	SF
۱۰۰	۱۰۰	۱۰۰	۵۴۶۰/۸۵۳	۲۳۱۵۴/۸۸	۵۴۶۰/۸۰۷	۵۳۱۴/۵۴۱	۵۴۶۰/۸۵۳
۲۰۰	۲۰۰	۲۰۰	۳۸۶۸/۴۱۹	۱۵۱۷۸/۴۹	۳۸۶۸/۳۵۹	۳۵۹۷/۷۷	۳۸۶۸/۴۱۹
۵۰۰	۵۰۰	۵۰۰	۲۳۹۳/۸۵۴	۸۷۲۳/۱۶۶	۲۳۹۳/۶۲۳	۵۳۱۴/۵۴۱	۲۳۹۳/۸۵۴

*نمادهای WA، Hart، FB، PML و SF به‌ترتیب بیانگر براوردگر تعدیل وزنی، هارتلی، فولر-بورمیستر، ماکسیمم درست‌نمایی‌نما و تک‌چارچوبی هستند.

از آن‌جایی که مقدار بهین $\hat{\theta}_H$ در براوردگر هارتلی به براورد ماتریس کوواریانس مجموع جامعه‌ای براوردشده بستگی دارد لذا این براوردگر $\sqrt{\text{EMSE}}$ بیش‌تری نسبت به سایر براوردگرها دارد. اما براوردگر فولر-بورمیستر با استفاده از اطلاعات اضافی (براورد اندازه‌ی جامعه‌ای) تا حد زیادی این مشکل را مرتفع می‌سازد.

جدول ۲ـ ریشه‌ی دوم میانگین توان دوم خطای تجربی براوردگرهای ۵گانه‌ی جامعه‌ی دوم

$n^{(۱)}$	$n^{(۲)}$	$n^{(۳)}$	WA*	Hart	FB	PML	SF
۱۰۰	۱۰۰	۱۰۰	۹۱۵۲/۴۰۶	۲۶۲۲۵/۰۲	۹۱۵۲/۳۵۳	۸۹۶۴/۶۷۸	۹۶۲۰/۵۴۹
۲۰۰	۲۰۰	۲۰۰	۶۳۱۷/۷۵	۱۸۴۳۰/۱۶	۶۳۱۳/۸۶۵	۶۲۶۲/۰۷۵	۶۵۴۷/۰۱۶
۵۰۰	۵۰۰	۵۰۰	۴۳۱۴/۷۲۸	۱۶۴۹۰/۹۷	۴۳۱۴/۶۹۷	۴۲۲۵/۰۶۷	۴۴۳۱/۹۳۵

همان‌طور که در جدول ۲ مشاهده می‌شود، روش ماکسیمم درست‌نمایی‌نما $\sqrt{\text{EMSE}}$ کوچک‌تری برای براورد مجموع جامعه‌ای داراست.

در عمل $\sqrt{\text{EMSE}}$ براوردگر فولر-بورمیستر بزرگ‌تر از $\sqrt{\text{EMSE}}$ ماکسیمم درست‌نمایی‌نما است. براوردگر فولر-بورمیستر به‌طور نظری بهین است اما برای محاسبه‌ی این براوردگر یک ماتریس کوواریانس ۱۴×۱۴ برای هر یک از سه نمونه باید براورد شود و تغییرپذیری اضافی در براورد کردن کوواریانس‌ها منجر به $\sqrt{\text{EMSE}}$ بزرگ‌تری برای روش فولر-بورمیستر می‌شود.

براوردگر هارتلی در آمارگیری سه‌چارچوبی عمل‌کرد ضعیفی را از خود نشان می‌دهد. در بعضی تکرارها، یک یا بیش‌تر مؤلفه‌های $\hat{\theta}_H$ منفی بودند، برای مثال براورد مجموع جامعه‌ای در حوزه‌ی abc به‌صورت $\hat{Y}_{abc} = -۰/۵\hat{Y}_{abc}^{A} + ۱/۲\hat{Y}_{abc}^{B} + ۰/۳\hat{Y}_{abc}^{C}$ است. گاهی اوقات ناپایداری براوردگر هارتلی سه‌چارچوبی با اندازه نمونه‌های کوچک موجب $\sqrt{\text{EMSE}}$ بیش‌تری نسبت به براوردگر میانگین‌گیری می‌شود. برای مثال، در جدول ۲،

$\sqrt{\text{EMSE}}$ براوردگر هارتلی برای براورد مجموع جامعه‌ای ۲۶۲۲۵/۰۲ است در حالی که $\sqrt{\text{EMSE}}$ روش تعدیل وزنی ۹۱۵۲/۴۰۶ است. البته با افزایش اندازه‌ی نمونه‌ای در هر یک از سه چارچوب براوردگر هارتلی پایدارتر خواهد شد. با این حال همچنان $\sqrt{\text{EMSE}}$ بزرگ‌تری نسبت به روش ماکسیمم درست‌نمایی‌نما داشته است.

۵- بحث و نتیجه‌گیری

ما در این مقاله چندین براوردگر نقطه‌ای از جمله تعدیل وزنی، هارتلی، فولر-بورمیستر، ماکسیمم درست‌نمایی‌نما و تک‌چارچوبی برای مجموع جامعه‌ای در آمارگیری‌های چندچارچوبی معرفی و درستی آن‌ها را با هم مقایسه کردیم.

براوردگر ماکسیمم درست‌نمایی‌نما به‌دلیل این‌که از وزن‌های یکسانی برای همه‌ی متغیرها استفاده می‌کند نسبت به براوردگرهای هارتلی و فولر-بورمیستر برتری دارد. دیگر براوردگرهای معرفی‌شده در آمارگیری سه‌چارچوبی عمل‌کردشان به‌خوبی براوردگر ماکسیمم درست‌نمایی‌نما نبود و به‌کارگیری آن‌ها در آمارگیری‌ها با بیش از دو چارچوب توصیه نمی‌شوند.

براوردگرهای هارتلی و فولر-بورمیستر ناپایدار هستند، زیرا حل نظام معادله‌های آن‌ها به استفاده از ماتریس کوواریانس‌های براوردشده‌ی بزرگ نیاز دارند.

براوردگر تک‌چارچوبی نیز عمل‌کرد ضعیفی نسبت به براوردگر ماکسیمم درست‌نمایی‌نما در بیش‌تر شبیه‌سازی‌ها داشته است. البته با استفاده از تعدیل وزن چنگک‌زنی یا رگرسیونی می‌توان به نتیجه بهتری دست یافت.

با این حال در همه‌ی شبیه‌سازی‌های صورت گرفته، براوردگر ماکسیمم درست‌نمایی‌نما کارایی بالاتری نسبت به سایر براوردگرها از خود نشان داده است. بنا بر این یک انتخاب مناسب برای براورد مجموع در آمارگیری‌های سه‌چارچوبی خواهد بود.

در بسیاری از موارد، آمارگیری‌های چندچارچوبی بدون افزایش هزینه‌ی گرداوری داده‌ها، درستی $\frac{1}{(\text{EMSE})}$ براوردهای مقدارهای جامعه‌ای را بهبود می‌بخشند. البته در انتخاب براوردگر و طرح نمونه‌گیری باید دقت کرد. برای مثال در جامعه‌های کمیاب با ترکیب چارچوب‌های ناقص، پوشش و درستی براوردها افزایش می‌یابد.

مرجع‌ها

[1] Bankier, M.D. (1986). Estimator Basedon Several Stratified Samples with Applications to Multiple Frame Surveys, *Journal of the American Statistical Association*, **81**, 1074–1079.

[2] Fuller, W.A. and Burmeister, L.F. (1972). Estimators for Samples Selected from Two Overlapping Frames, in Proceeding of the Statistics Section, *American Statistical Association*, 245–249.

[3] Haines, D.E. and Pollock, K.H. (1998). Combining Multiple Frames to Estimate Population Size and Totals, *Survey Methodology*, **24**, 79–88.

[4] Hajek, J. (1971). Comment on an Essay on the Logical Foundations of Survey Sampling, Part one, In V. Godambe and D. Sprott (Eds.), Foundations of Statistical Inference. Holt, Rinehart and Winston Toronto.

[5] Hartley, H.O. (1962). Multiple Frame Surveys, in Proceedings of the Social Statistics Section, *American Statistical Association*, 203–206.

[6] Hartley, H.O. (1974). Multiple Frame Methodology and Selected Applications, *Sankhya*, Ser. C, **36**, 99–118.

[7] Iachan, R. and Dennis, M.L. (1993). Multiple Frame Approach to Sampling the Homeless and Transient Population, *Journal of official Statistics*, **9**, 747–764.

[8] Kalton, G. and Anderson, D.W. (1986). Sampling Rare Populations, *Journal of the Royal Statistical Society*, Ser. A, **149**, 65–82.

[9] Lohr, S.L. and Rao, J.N.K. (2000). Inference in Dual Frame Surveys, *Journal of the American Statistical Association*, **95**, 271–280.

[10] Lohr, S.L. and Rao, J.N.K. (2006). Estimation in Multiple-Frame Surveys, *Journal of the American Statistical Association*, **101**, 1019–1030.

[11] Lohr, S.L. (2009). Multiple Frame Surveys, In Handbook of Statistics, Sample Surveys: Design, Methods and Applications, (Eds.,

D. Pfeffermann and C. R. Rao). Amsterdam: North Holland, **29**, 71-88.

[12] Lohr, S.L. (2011). Alternative Survey Sample Designs: Sampling with Multiple Overlapping Frames, *Survey Methodology*, **37**, 197-213.

[13] Lohr, S.L. and Lu, Y. (2010). Gross Flow Estimation in Dual Frame Surveys, *Survey Methodology*, **36**, 13-22.

[14] Mecatti, F. (2007). A Singl Frame Multiplicity Estimator for Multiple Frame Surveys, *Survey Methodology*, **33**, 151-157.

[15] Rao, J.N.K. and Wu, C. (2010). Pseudo-Empirical Likelihood Inference for Dual Frame Surveys, *Journal of the American Statistical Association*, **105**, 1494-1503.

[16] Skinner, C.J. (1991). On the Efficiency of Raking Ratio Estimation for Multiple Frame Surveys, *Journal of the American Statistical Association*, **86**, 779-784.

[17] Skinner, C.J. and Rao, J.N.K. (1996). Estimation in Dual Frame Surveys With Complex Designs, *Journal of the American Statistical Association*, **91**, 349-356.

طیبه چگینی
فوق لیسانس آمار
تهران، خیابان شهید بهشتی، نبش احمد قصیر، دانشگاه علامه طباطبایی، دانشکده‌ی اقتصاد، گروه آمار.
رایانشانی: taiebehchegini@yahoo.com

حمیدرضا نواب‌پور
دکتری آمار
تهران، خیابان شهید بهشتی، نبش احمد قصیر، دانشگاه علامه طباطبایی، دانشکده‌ی اقتصاد، گروه آمار.
رایانشانی: h.navvabpour@srtc.ac.ir