دانشگاه صنعتی اصفهان

دانشکده علوم ریاضی

آنتروپی طرح‌های نمونه‌گیری با احتمال نابرابر

پایان‌نامه کارشناسی ارشد (آماراقتصادی—اجتماعی)

فهیمه مسیحی بیدگلی

استاد راهنما

دکتر سروش علیمرادی

۱۳۹۰/۸/۲۱

دانشگاه صنعتی اصفهان

دانشکده علوم ریاضی

پایان‌نامه کارشناسی ارشد (آماراقتصادی—اجتماعی)خانم فهیمه مسیحی بیدگلی

تحت عنوان

آنتروپی طرح‌های نمونه‌گیری با احتمال نابرابر

در تاریخ ۱۳۹۰ توسط کمیته تخصصی زیر مورد بررسی و تصویب نهائی قرار گرفت.

۱— استاد راهنمای پایان‌نامه	دکتر سروش علیمرادی
۲— استاد مشاور پایان‌نامه	دکتر علی رجالی
۳— استاد داور ۱	دکتر محمد محمدی (دانشگاه اصفهان)
۴— استاد داور ۲	دکتر علی زینل همدانی

سرپرست تحصیلات تکمیلی دانشکده — دکتر اعظم اعتماد

فهرست مطالب

چکیده:

ساده‌ترین روش نمونه‌گیری احتمالی، روش نمونه‌گیری تصادفی ساده است که در آن احتمال قرار گرفتن هر واحد جامعه در نمونه یکسان است. اما هنگامی که اندازه‌ی واحدهای جامعه تغییرپذیری قابل توجهی داشته باشند، این روش مناسب نخواهد بود. در چنین شرایطی از نمونه‌گیری با احتمال نابرابر استفاده می‌شود.

برای نمونه‌گیری با احتمال نابرابر طرح‌های بسیاری نظیر نمونه‌گیری پواسون، پواسون شرطی، سامفورد، پارتو، روش خرد کردن و ... وجود دارد. می‌توان با استفاده از معیاری به نام آنتروپی، طرح‌های مختلف را با یکدیگر مقایسه نمود. آنتروپی یک طرح نمونه‌گیری میزان تصادفی بودن و گستردگی آن طرح را نشان می‌دهد. همچنین برای مقایسه می‌توان فاصله‌ی طرح‌های مختلف را محاسبه کرد.

در این پایان‌نامه آنتروپی طرح‌های نمونه‌گیری با احتمال نابرابر محاسبه شده و نشان داده می‌شود که چندین طرح در داشتن حداکثر آنتروپی به یکدیگر نزدیک هستند. همچنین نزدیک بودن تابع احتمال این طرح‌ها با استفاده از فاصله‌ی هلینگر تأیید شده است.

رده‌بندی موضوعی: ۶۲D۰۵.

کلمات کلیدی: آنتروپی، طرح نمونه‌گیری، فاصله‌ی هلینگر، نمونه‌گیری با احتمال نابرابر

فصل ۱

مقدمه و تاریخچه

به‌طور معمول آگاهی افراد درباره‌ی مشخصه‌های هر جامعه از گذشته‌های دور از طریق نمونه‌گیری فراهم شده است. هدف روش‌های نمونه‌گیری، تهیه‌ی اطلاعات از جامعه با مطالعه‌ی تنها زیرمجموعه‌ای از آن به نام نمونه است. در واقع نمونه‌گیری فرآیند انتخاب واحدها از جامعه است به‌طوری‌که به کمک آن‌ها بتوان مقادیر نامعلوم جامعه را به‌شکل مطلوب برآورد کرد. بنابراین یکی از مسائل حائز اهمیت در نمونه‌گیری، تطابق مجموعه‌ی نمونه با جامعه‌ی کل است.

یک نمونه باید معرف و نماینده‌ی مناسبی برای جامعه‌ی هدف باشد. در تعریف نمایندگی اغلب گفته می‌شود، یک نمونه در صورتی نماینده‌ی مناسبی برای جامعه است که نسخه‌ی کوچک و یا مدل کاهش‌یافته‌ی جامعه بوده و طبقات با همان نسبتی که در جامعه‌اند، باید در نمونه نیز ظاهر شوند. اما این نظریه‌ی رایج بین عموم افراد، در بعضی مواقع نادرست است. برای مثال فرض کنید هدف، برآورد تولید آهن در یک کشور باشد. آهن تولیدشده از یک سو به‌وسیله‌ی دو شرکت بزرگ فولاد با چندین هزار کارگر و از سوی دیگر به‌وسیله‌ی چندین شرکت صنعتی کوچک با کمتر از ۵۰ کارگر در هر یک فراهم می‌شود. در این صورت روش نمونه‌گیری که در آن شانس انتخاب هر عضو جمعیت با دیگری برابر باشد، مناسب نخواهد بود. برای برآورد دقیق باید از تولیدهای دو شرکت بزرگ در نمونه استفاده شود. این مثال ساده نشان‌دهنده‌ی استفاده از نمونه‌گیری با احتمال نابرابر است [۵۷].

نمونه‌گیری با احتمال نابرابر از جمله روش‌های نمونه‌گیری احتمالی است. به‌طور کلی دو نوع روش نمونه‌گیری وجود دارد که به روش نمونه‌گیری احتمالی و روش نمونه‌گیری غیر احتمالی معروف‌اند.

نمونه‌گیری احتمالی که در آن هر واحد نمونه با احتمالی مشخص از جامعه استخراج می‌شود، از سال ۱۹۴۰ به بعد به‌طور چشمگیری گسترش یافته و تقریباً در همه‌ی کاربردها، جانشین نمونه‌گیری غیر احتمالی شده است. این نوع نمونه‌گیری که توسط دمینگ[۱] [۱۵] در سال ۱۹۵۰ به این نام خوانده شد، امروزه در صنعت، تجارت، نظرخواهی‌های عمومی، بازاریابی، کنترل کیفیت و بیشتر زمینه‌های پژوهشی به‌کار می‌رود[۶۲].

ساده‌ترین روش نمونه‌گیری احتمالی، روش نمونه‌گیری تصادفی ساده است. در این روش تمام نمونه‌های ممکن از شانسی برابر برای انتخاب شدن برخوردارند و لذا احتمال قرار گرفتن هر واحد جامعه در نمونه یکسان است. اما هنگامی که اندازه‌ی واحدها تغییرپذیری قابل توجهی دارند، این روش مناسب نخواهد بود. زیرا در نمونه‌گیری تصادفی ساده، اهمیت اندازه‌ی واحدها در نظر گرفته نمی‌شود. در چنین شرایطی انتخاب واحدها با احتمال نابرابر، برآوردگرهای کاراتری نسبت به نمونه‌گیری با احتمال برابر ایجاد خواهد کرد[۵۴].

استفاده از نمونه‌گیری با احتمال نابرابر(UPS)[۲] ، اولین بار توسط هنسن و هرویتز[۳] [۲۶] در سال ۱۹۴۳ پیشنهاد شد. آن‌ها در یک طرح نمونه‌گیری دومرحله‌ای، از نمونه‌گیری با احتمال نابرابر به روش با جایگذاری استفاده کردند. شیوه‌ی نمونه‌گیری هنسن و هرویتز به‌عنوان نمونه‌گیری متناسب با اندازه و با جایگذاری($PPSWR$)[۴] شناخته شد. شیوه‌های طرح با جایگذاری دارای کارایی کمتری نسبت به طرح‌های بدون جایگذاری هستند. از این‌رو سعی در ایجاد شیوه‌های نمونه‌گیری با احتمال نابرابر و بدون جایگذاری شد. مادو[5] [۳۵] در سال ۱۹۴۹ استفاده از نمونه‌گیری سیستماتیک با احتمال نابرابر را به‌طوری‌که واحدها بیشتر از یک بار انتخاب نشوند، پیشنهاد کرد.

نارین[۶] [۳۹] در سال ۱۹۵۱ و هارویتز و تامپسون[۷] [۳۰] در سال ۱۹۵۲، نظریه‌ی نمونه‌گیری با احتمال نابرابر را برای حالت بدون جایگذاری توسعه دادند. طرح‌های نمونه‌گیری با احتمال نابرابر و بدون جایگذاری به سرعت توسط محققین مختلف مورد بررسی قرار گرفت. اما بیشتر روش‌های پیشنهادی، محدود به اندازه‌ی نمونه‌ی برابر با ۲ بود. در طرح‌هایی که برای اندازه‌ی نمونه‌ای بیشتر، گسترش یافت محاسبه‌ی احتمال‌های شمول مرتبه‌ی دوم پیچیده و گاه با افزایش نمونه غیر عملی بود.

بریور و حنیف[۸] [۱۱] در سال ۱۹۸۳ فهرستی از ۵۰ روش نمونه‌گیری با احتمال نابرابر و بدون جایگذاری

[۱] Deming, W.E.
[۲] Unequal Probability Sampling
[۳] Hansen, M. H., Hurwitz, W.N.
[۴] Probability Proportional to Size With Replacement
[5] Madow, W.G.
[۶] Narain, R.D.
[۷] Horvitz, D.G., Thompson, D.J.
[۸] Brewer, K. R. W., Hanif, M.

را بر اساس تاریخ انتشار ارائه کردند. روش چائو[۹] [۱۳]، سانتر[۱۰] [۵۵] و[۵۶]، نمونه‌گیری ترتیبی[۱۱] و نمونه‌گیری خرد کردن[۱۲] از جمله روش‌های مهمی هستند که تکمیل‌کننده‌ی نمونه‌گیری با احتمال نابرابر بودند.

نمونه‌گیری ترتیبی یک کلاس جدید و مفید از طرح‌های نمونه‌گیری با احتمال نابرابر است که توسط رسن[۱۳] [۴۸] و [۴۹] در سال ۱۹۹۷ معرفی شد. این طرح‌ها در اجرا بسیار ساده، اما توصیف آن‌ها از نظر احتمالی مشکل است. دویل و تیل[۱۴] [۱۸] در سال ۱۹۹۸ طرح‌های جدیدی از نمونه‌گیری با احتمال نابرابر را با استفاده از روش خرد کردن ارائه کردند. در واقع روش خرد کردن کلاسی از روش نمونه‌گیری با احتمال نابرابر و بدون جایگذاری است که سایر طرح‌های بدون جایگذاری را می‌توان به‌وسیله‌ی آن ایجاد کرد. تیل[۱۵] [۵۷] در سال ۲۰۰۶ نمونه‌گیری به روش خرد کردن، طرح‌های نمونه‌گیری نمایی[۱۶] با احتمال نابرابر و نمونه‌گیری متعادل[۱۷] با احتمال نابرابر را با جزئیات کامل مورد بررسی قرار داد. کتاب وی پس از اثر بریور و حنیف[۱۱]، کتابی جامع در زمینه‌ی نمونه‌گیری با احتمال نابرابر است.

همان‌طور که ذکر شد، برای نمونه‌گیری با احتمال نابرابر طرح‌های بسیاری می‌تواند وجود داشته باشد. سؤالی که می‌توان مطرح کرد، چگونگی مقایسه‌ی طرح‌های مختلف با یکدیگر است.

یک طرح نمونه‌گیری مناسب، باید در اجرا آسان و کارا باشد. از دیگر ویژگی‌های یک طرح خوب، دقیق بودن آن است. در یک طرح دقیق احتمال‌های شمول واقعی با احتمال‌های شمول از پیش تعیین‌شده یکسان است. داشتن اطلاعات کافی درباره‌ی جامعه می‌تواند در انتخاب یک طرح نمونه‌گیری مناسب مؤثر باشد. اما زمانی که هیچ نوع اطلاعاتی در مورد جامعه وجود نداشته و امکان دارد فرض تناسب بین متغیر هدف و متغیر کمکی به خوبی برقرار نباشد، ویژگی مهمی که از یک طرح می‌تواند مطرح باشد، آنتروپی[۱۸] و دارا بودن سطح بالایی از تصادفی بودن است.

مفهوم آنتروپی برای اولین بار در اواسط سده‌ی نوزدهم توسط فیزیکدان آلمانی، رودلف کلوسیوس[۱۹] مطرح شد. تا قبل از مقاله‌ای که در سال ۱۹۴۸ توسط شانون[۲۰] [۵۳] منتشر شد، آنتروپی تنها در فیزیک

---

[۹] Chao, M.T.
[۱۰] Sunter, A.B.
[۱۱] Order Sampling
[۱۲] Splitting sampling
[۱۳] Rosen, B.
[۱۴] Deville, J.-C., Tille, Y.
[۱۵] Tille, Y.
[۱۶] Exponential sampling design
[۱۷] Balanced sampling design
[۱۸] Entropy
[۱۹] Clausius, R.
[۲۰] Shannon, C.E.

و مکانیک آماری، برای کمی‌سازی مفاهیمی همچون عدم قطعیت و بی‌نظمی به‌کار برده می‌شد. شانون در سال ۱۹۴۸، آنتروپی و معیار اطلاع را با علم آمار و احتمال پیوند داد که بدین صورت آنتروپی جایگاه خود را در بین سایر علوم تثبیت کرد.

آنتروپی توسط هاجک[۲۱] [۲۸] در سال ۱۹۸۱ در نمونه‌گیری از یک جمعیت متناهی مورد استفاده قرار گرفت. هاجک نشان داد که در کلاس طرح‌های نمونه‌گیری با احتمال نابرابر و بدون جایگذاری با اندازه‌ی نمونه‌ی ثابت و بردار احتمال شمول یکسان، طرح پواسون شرطی تعدیل‌شده(ACP)[۲۲] دارای حداکثر آنتروپی است. همچنین برای طرح‌های نمونه‌گیری با احتمال برابر و با اندازه‌ی ثابت، نمونه‌گیری تصادفی ساده(SRS)[۲۳] از حداکثر میزان تصادفی بودن برخوردار است.

گرافسترم[۲۴] [۲۲] در سال ۲۰۰۹ آنتروپی چندین طرح نمونه‌گیری با احتمال نابرابر را مورد بررسی قرار داد و به این نتیجه رسید که چندین طرح در داشتن حداکثر میزان آنتروپی به یکدیگر نزدیک هستند.

در این پایان‌نامه ابتدا نمونه‌گیری با احتمال نابرابر معرفی شده و سپس طرح‌های مختلف با استفاده از معیار آنتروپی و فاصله‌ی هلینگر[۲۵] مقایسه می‌شوند. در این فصل، تاریخچه و مقدمه‌ای از نمونه‌گیری با احتمال نابرابر آورده شد. در ادامه، در فصل دوم برخی از تعاریف و مفاهیم اولیه‌ی مورد نیاز در فصل‌های بعدی توضیح داده می‌شود. در فصل سوم به معرفی نمونه‌گیری با احتمال نابرابر در انواع کلاس‌ها و برآوردگرهای متناظر خواهیم پرداخت. در فصل چهارم چندین طرح نمونه‌گیری متناسب با اندازه و بدون جایگذاری(πps)[۲۶] مورد بررسی قرار می‌گیرد. فصل پنجم به معرفی آنتروپی، انواع برآورد و کاربردهایی از آن اختصاص می‌یابد. در فصل ششم به محاسبه‌ی آنتروپی و فاصله‌ی هلینگر برای مقایسه‌ی طرح‌های نمونه‌گیری با احتمال نابرابر پرداخته و در انتها در فصل هفتم نتیجه‌گیری بیان می‌شود.

---

[۲۱] Hajek, J.
[۲۲] Adjusted Conditional Poisson
[۲۳] Simple Random Sampling
[۲۴] Grafstrom, A.
[۲۵] Hellinger distance
[۲۶] Inclusion probability (π) proportional to size

فصل ۲

تعاریف و مفاهیم اولیه

از آنجا که آگاهی از مفاهیم اساسی و تعاریف مربوطه لازمه‌ی مطالعه در هر زمینه است، در این فصل ابتدا به معرفی برخی از نمادها و سپس تعاریف و اصطلاحاتی می‌پردازیم که در فصل‌های بعدی مورد استفاده قرار خواهند گرفت. در ادامه نیز انواع طرح‌ها و الگوریتم‌های نمونه‌گیری آورده می‌شود.
لازم به ذکر است در این پایان‌نامه، متغیر هدف در جامعه‌ی $\mathcal{U}$ با نماد y و مقدار آن برای هر $k \in \mathcal{U}$ با y_k نمایش داده می‌شود. جمع متغیر هدف نیز $Y = \sum_{k\in\mathcal{U}} y_k$ است. همچنین فرض می‌شود که به‌ازای هر واحد k در جامعه، اطلاعات متغیر کمکی z_k وجود داشته و متغیر y به‌طور تقریبی با متغیر کمکی z متناسب است. سایر اختصارات و نمادهای به‌کار رفته در پایان‌نامه در پیوست ۱ آورده شده است.

۲–۱ تعاریف

مطالب این بخش اکثراً بر اساس مراجع [۵۰] و [۵۷] است.

تعریف ۱.۲ **جامعه‌ی متناهی**[۱] : یک جامعه‌ی متناهی مجموعه‌ای از N واحد به‌صورت $\mathcal{U} = \{\mathcal{U}_1, \ldots, \mathcal{U}_k, \ldots, \mathcal{U}_N\}$ است. مقدار N، اندازه‌ی جامعه نامیده می‌شود. می‌توان به هر واحد جامعه عددی از ۱ تا N به‌عنوان برچسب اختصاص داد و جامعه‌ی $\mathcal{U}$ را به‌صورت زیر نمایش داد:

$$\mathcal{U} = \{1, \ldots, k, \ldots, N\}$$

تعریف ۲.۲ **نمونه**[۲] : عناصری از جامعه که توسط محقق انتخاب و مشخصات آن‌ها اندازه‌گیری می‌شوند، تشکیل یک نمونه می‌دهند. معمولاً نمونه با نماد s نمایش داده می‌شود. در نظریه‌ی نمونه‌گیری دو تعریف برای اصطلاح نمونه وجود دارد.

الف) نمونه‌ی با جایگذاری که به‌صورت یک مجموعه‌ی مرتب از جامعه‌ی $\mathcal{U}$ در نظر گرفته می‌شود و با نماد زیر نمایش می‌دهند:

$$\{k_1, k_2, \ldots, k_{n(s)}\}$$

که در آن $k_i \in \mathcal{U}$، عنصر انتخاب‌شده در i–اُمین ($(i = 1, 2, \ldots, n(s))$) استخراج است. $n(s)$ تعداد عناصر نمونه‌ی s است و آن را اندازه‌ی نمونه می‌نامند. در صورت ثابت بودن اندازه‌ی نمونه، به‌طور ساده می‌توان از نماد n استفاده کرد. در نمونه‌ی با جایگذاری واحدهای انتخاب‌شده لزوماً متفاوت نیستند.

ب) نمونه‌ی بدون جایگذاری، یک زیرمجموعه‌ی غیر تهی از $\mathcal{U}$ با n عنصر است. در این حالت واحدهای انتخاب‌شده الزاماً متفاوت هستند.

همچنین در یک سیستم نام‌گذاری دیگر، نمونه را بر اساس ترتیب استخراج به دو دسته‌ی نمونه‌ی مرتب و نمونه‌ی نامرتب تقسیم‌بندی می‌کنند. اگر در فرآیند نمونه‌گیری ترتیب استخراج واحدها در نظر گرفته شود، نمونه‌ی مرتب و در غیر این صورت، نمونه‌ی نامرتب به‌وجود می‌آید.

تعریف ۳.۲ **تابع نشانگر عضویت نمونه**[۳] : انتخاب واحد k در نمونه، متغیری تصادفی است که با I_k نمایش می‌دهند و به‌صورت زیر تعریف می‌شود:

$$I_k = \begin{cases} 1 & k \in S \\ 0 & \text{در غیر این صورت.} \end{cases}$$

---

[۱] Finite population
[۲] Sample
[۳] Sample membership indicator function

برخی آماردانان روش به‌کار رفته برای نمونه‌گیری و برآورد پارامترهای جامعه را طرح نمونه‌گیری می‌نامند. امروزه برای بیان دقیق آن، مفاهیم زیر مورد نیاز است:

تعریف ۴.۲ بردار طرح[۴] : اگر I_k^* نشان‌دهنده‌ی تعداد قرار گرفتن واحد k–اُم جامعه در نمونه باشد، آن‌گاه بردار تصادفی $\mathbf{I} = (I_1^*, \ldots, I_N^*)$ را بردار طرح می‌نامند. توزیع بردار طرح $\mathbf{I}$، توزیع طرح نامیده می‌شود.

در یک نمونه‌گیری بدون جایگذاری، I_k^* معادل با I_k (تابع نشانگر عضویت) خواهد بود. I_k متغیر تصادفی برنولی و I_k^* دارای توزیع دوجمله‌ای است. می‌توان بردار طرح $\mathbf{I}$ را به‌عنوان نمونه‌ی تصادفی در نظر گرفت. در این صورت تحققی از بردار طرح را با $\mathbf{x}$ نشان می‌دهند که یک نمونه‌ی مشاهده‌شده خواهد بود. هر $\mathbf{x}$ به‌طور یکتا با s (زیرمجموعه‌ای از جامعه) مطابق است.

بیان نمونه بر اساس بردار طرح

الف) نمونه‌ی بدون جایگذاری

یک نمونه‌ی بدون جایگذاری را به‌عنوان برداری از نشانگر شمول به‌صورت زیر نمایش می‌دهند:

$$\mathbf{x} = (x_1, \ldots, x_k, \ldots, x_N) \in \{0, 1\}^N$$

که در آن

$$x_k = \begin{cases} 1 & \text{اگر واحد } k\text{–اُم در نمونه باشد} \\ 0 & \text{در غیر این صورت.} \end{cases}$$

ب) نمونه‌ی با جایگذاری

یک نمونه‌ی با جایگذاری با بردار زیر مشخص می‌شود:

$$\mathbf{x} = (x_1, \ldots, x_k, \ldots, x_N) \in \{0, 1, 2, \ldots\}^N$$

که در آن x_k تعداد دفعاتی است که واحد k–اُم در نمونه قرار می‌گیرد.

تعریف ۵.۲ تکیه‌گاه طرح[۵] : تکیه‌گاه $\mathcal{Q}$، مجموعه‌ای از تمام نمونه‌های ممکن جامعه برای یک طرح است. تعداد نمونه‌های ممکن را اندازه‌ی تکیه‌گاه نامیده و آن را با $card(\mathcal{Q})$ نمایش می‌دهند. تکیه‌گاه $\mathcal{Q}$ متقارن گفته می‌شود، اگر برای هر $s \in \mathcal{Q}$ تمام جایگشت‌های مختصات s در $\mathcal{Q}$ باشد.

[۴] Design vector
[۵] Support of design

انواع تکیه‌گاه‌های متقارن به‌صورت زیر است:
الف) تکیه‌گاه متقارن بدون جایگذاری که به‌صورت $\mathcal{S}=\{0,1\}^N$ بیان می‌شود و $card(\mathcal{S})=2^N$ است.
ب) تکیه‌گاه متقارن بدون جایگذاری با اندازه‌ی نمونه‌ی ثابت n که عبارت است از:

$$\mathcal{S}_n=\{\mathbf{x}\in\mathcal{S}\mid\sum_{k\in\mathcal{U}}x_k=n\}$$

و $card(\mathcal{S}_n)=\binom{N}{n}$ است.
پ) تکیه‌گاه متقارن با جایگذاری $\mathcal{R}=\mathbb{N}^N$ است که $\mathbb{N}$ مجموعه‌ی اعداد طبیعی است. این تکیه‌گاه، شمارش پذیر نامتناهی بوده و $card(\mathcal{R})=\infty$ است.
ت) تکیه‌گاه متقارن با جایگذاری با اندازه‌ی نمونه‌ی ثابت n که به‌صورت زیر است:

$$\mathcal{R}_n=\{\mathbf{x}\in\mathcal{R}\mid\sum_{k\in\mathcal{U}}x_k=n\}$$

اندازه‌ی تکیه‌گاه $\mathcal{R}_n$، $card(\mathcal{R}_n)=\binom{N+n-1}{n}$ است.

تعریف ۶.۲ طرح نمونه‌گیری[۶] : یک طرح نمونه‌گیری p با تکیه‌گاه $\mathcal{Q}$ یک توزیع احتمال چندمتغیره روی $\mathcal{Q}$ است. بنابراین p یک تابع از تکیه‌گاه $\mathcal{Q}$ به $[0,1)$ است به‌طوری‌که،
۱– به ازای هر نمونه‌ی مفروض $s\in\mathcal{Q}$ ، $p(s)>0$.
۲– جمع توزیع احتمال روی تمام نمونه‌های ممکن یک شود. به‌عبارت دیگر، $\sum_{s\in\mathcal{Q}}p(s)=1$.
به p تابع احتمال نیز گفته می‌شود که از طریق توزیع بردار طرح، $Pr(\mathbf{I}=\mathbf{x})$، قابل محاسبه است.

یک طرح نمونه‌گیری با تکیه‌گاه $\mathcal{Q}$ در صورتی که $\mathcal{Q}\subset\mathcal{S}$ باشد، بدون جایگذاری نامیده می‌شود. اگر $\mathcal{Q}\backslash\mathcal{S}$ (به‌جز $\mathcal{Q}$) غیر تهی باشد، با جایگذاری است و اگر $\mathcal{Q}\subset\mathcal{R}_n$ باشد، طرح با اندازه‌ی ثابت n خواهد بود.

تعریف ۷.۲ نمونه‌ی تصادفی : نمونه‌ی تصادفی S با طرح نمونه‌گیری p یک مجموعه‌ی تصادفی است، به‌طوری‌که:

$$Pr(\mathrm{S}=s)=p(s)\quad\forall\, s\in\mathcal{Q}$$

تعریف ۸.۲ امید ریاضی و واریانس نمونه‌ی تصادفی[۷] : امید ریاضی یک نمونه‌ی تصادفی S عبارت است از:

$$\boldsymbol{\mu}=E(\mathrm{S})=\sum_{\mathbf{x}\in\mathcal{Q}}\mathbf{x}p(\mathbf{x})$$

[۶] Sampling design
[۷] Expectation and variance of a random sample

به $\boldsymbol{\mu}$ بردار میانگین طرح گفته می‌شود. ماتریس واریانس–کواریانس نمونه‌ی تصادفی نیز به‌صورت زیر تعریف می‌شود:

$$\boldsymbol{\Sigma} = [\Sigma_{kl}] = Var(\mathrm{S}) = \sum_{\mathbf{x}\in\mathcal{Q}} p(\mathbf{x})(\mathbf{x}-\boldsymbol{\mu})'(\mathbf{x}-\boldsymbol{\mu}) = [\mu_{kl} - \mu_k\mu_l]$$

که در آن $\mu_{kl} = \sum_{\mathbf{x}\in\mathcal{Q}} x_k x_l p(\mathbf{x})$، امید ریاضی توأم نمونه‌ی S است.

تعریف ۹.۲ احتمال شمول مرتبه‌ی اول[۸] : احتمال آن که واحد k–اُم در نمونه‌ی تصادفی قرار گیرد. به‌عبارت دیگر احتمال شمول برای واحد k–اُم با جمع کردن احتمال نمونه‌های ممکنی که شامل k–اُمین واحد می‌شوند، به‌دست می آید و با نماد π_k نمایش داده می‌شود. در این صورت، برای $k = 1, 2, \ldots, N$ داریم:

$$\pi_k = Pr(I_k = 1) = E(I_k) = \sum_{\mathbf{x}\in\mathcal{Q}} x_k p(\mathbf{x}) = \sum_{s\ni k} p(s)$$

بردار $\boldsymbol{\pi}=(\pi_1, \ldots, \pi_k, \ldots, \pi_N)$ بردار احتمال شمول نامیده می‌شود. در صورتی که طرح نمونه‌گیری بدون جایگذاری باشد، $\boldsymbol{\pi} = \boldsymbol{\mu}$ خواهد بود.

مثال ۱.۲ اگر جامعه‌ی $\mathcal{U} = \{1, 2, 3, 4\}$ در نظر گرفته شود، ٦ نمونه‌ی ممکن

$$s_1 = \{1, 2\}, s_2 = \{1, 3\}, s_3 = \{1, 4\}, s_4 = \{2, 3\}, s_5 = \{2, 4\}, s_6 = \{3, 4\}$$

نمونه‌های بدون جایگذاری با اندازه‌ی ۲ هستند. اگر یک طرح نمونه‌گیری با احتمال نابرابر وجود داشته باشد، نمونه‌ها دارای احتمال‌های متفاوتی خواهند بود. نمونه‌های s_1 تا s_3 با احتمال $\frac{1}{9}$ و نمونه‌های s_4 تا s_6 با احتمال $\frac{2}{9}$ انتخاب می‌شوند. در این صورت احتمال شمول برای واحد ۱، $\pi_1 = \sum_{s\ni 1} p(s) = 3 \times \frac{1}{9} = \frac{1}{3}$ و برای واحدهای $k = 2, 3, 4$، $\pi_k = \sum_{s\ni k} p(s) = \frac{1}{9} + \frac{2}{9} + \frac{2}{9} = \frac{5}{9}$ خواهد بود[۲۴].

تعریف ۱۰.۲ احتمال شمول مرتبه‌ی دوم[۹] : احتمال انتخاب توأم واحدهای k و l در نمونه‌ی تصادفی است و با جمع کردن احتمال نمونه‌هایی که شامل واحدهای k و l به‌طور هم‌زمان هستند، حاصل می‌شود. احتمال شمول مرتبه‌ی دوم را با π_{kl} نمایش می‌دهند و به‌ازای $k, l = 1, 2, \ldots, N$،

$$\pi_{kl} = Pr(I_k = 1, I_l = 1) = E(I_k I_l) = \sum_{\mathbf{x}\in\mathcal{Q}} x_k x_l p(\mathbf{x}) = \sum_{s\ni k,l} p(s)$$

[۸] First-order inclusion probability

[۹] Second-order inclusion probability

$\mathbf{\Pi} = [\pi_{kl}]$ را ماتریس احتمال‌های شمول توأم (مرتبه‌ی دوم) می‌نامند. اگر $\mathbf{\Delta} = \mathbf{\Pi} - \boldsymbol{\pi}\boldsymbol{\pi}'$ تعریف کنیم، در صورتی که طرح بدون جایگذاری باشد، $\mathbf{\Delta} = \mathbf{\Sigma}$.

برای یک طرح نمونه‌گیری مفروض p، تابع I_k دارای خصوصیات زیر است:

- $E(I_k) = \pi_k, \qquad k = 1, \dots, N$
- $Var(I_k) = \pi_k(1 - \pi_k), \qquad k = 1, \dots, N$
- $Cov(I_k, I_l) = \pi_{kl} - \pi_k\pi_l = \Delta_{kl}, \qquad k \neq l$

اگر طرح نمونه‌گیری p دارای اندازه‌ی ثابت n باشد، آن‌گاه

- $\sum_{k \in \mathcal{U}} \pi_k = n$
- $\sum_{k \in \mathcal{U}} \sum_{l \neq k \in \mathcal{U}} \pi_{kl} = n(n-1)$
- $\sum_{l \in \mathcal{U}} \pi_{kl} = (n-1)\pi_k, \qquad l \neq k$

تعریف ۱۱.۲ **الگوریتم نمونه‌گیری**[۱۰] : یک الگوریتم نمونه‌گیری، شیوه‌ای است که برای انتخاب یک نمونه‌ی تصادفی به‌کار می‌رود. هدف الگوریتم نمونه‌گیری انتخاب یک نمونه، صرف‌نظر از بررسی تمام نمونه‌های ممکن است.

تعریف ۱۲.۲ **روش استخراج واحد به واحد**[۱۱] : یک روش نمونه‌گیری با اندازه‌ی نمونه‌ی ثابت n، روش با استخراج واحد به واحد گفته می‌شود، هرگاه در هر یک از n مرحله‌ی نمونه‌گیری، یک واحد به‌طور قطع در نمونه انتخاب شود.

تعریف ۱۳.۲ **روش فهرست—دنباله‌ای**[۱۲] : یک روش فهرست—دنباله‌ای، شیوه‌ای است که برای فهرستی از واحدها که طبق یک ترتیب خاص قرار گرفته‌اند، به‌کار می‌رود. می‌توان این ترتیب را با $1, 2, \dots, N$ نمایش داد.

به‌عنوان مثال فرض کنید عناصر داخل فهرست در یک ترتیب مشخص $(k = 1, \dots, N)$ قرار گیرند. اگر $u_1, \dots, u_N$ ، تعداد N مشاهده از توزیع یکنواخت $U(0, 1)$ بوده و همچنین $0 < \pi \leq 1$، آن‌گاه در صورتی که $u_k < \pi$، واحد k—اُم انتخاب می‌شود. بنابراین نشانگرهای عضویت نمونه $I_1, \dots, I_N$ مستقل و هم‌توزیع خواهند بود.

$$Pr(I_k = 1) = \pi \qquad , \qquad Pr(I_k = 0) = 1 - \pi$$

[۱۰] Sampling algorithm
[۱۱] Draw by draw method
[۱۲] List-sequential method

احتمال شمول برای همه‌ی واحدها برابر π و احتمال شمول توأم $\pi^{۲}$ است. در این نوع نمونه‌گیری اندازه‌ی نمونه‌ی $n(s)$ متغیر تصادفی است که دارای توزیع $B(N,\pi)$ (دوجمله‌ای با پارامترهای N و π) است.
تکیه‌گاه نمونه‌گیری فوق، S بوده و طرح نمونه‌گیری عبارت است از:

$$p(s) = \pi^{n(s)}(۱-\pi)^{N-n(s)}$$

که به آن طرح نمونه‌گیری برنولی[۱۳] گفته می‌شود.
نمونه‌گیری برنولی از جمله نمونه‌گیری‌های با احتمال برابر و جزء خانواده‌ی طرح‌های ساده است. در صورت تبدیل تکیه‌گاه به S_n، نمونه‌گیری تصادفی ساده حاصل می‌شود. همچنین تعمیم نمونه‌گیری برنولی منجر به نمونه‌گیری پواسون[۱۴] می‌شود. طرح نمونه‌گیری پواسون از جمله طرح‌های نمونه‌گیری با احتمال نابرابر است که در فصل سوم مورد بررسی قرار خواهد گرفت.

تعریف ۱۴.۲ روش رد—پذیرش $(A-R)$[۱۵] : در این روش نمونه‌هایی از یک طرح در صورتی که در شرط خاصی صدق کنند، می‌توانند به‌عنوان نمونه‌ی طرح دیگر پذیرفته شوند. به‌عبارت دیگر، اگر $p_۱$ و $p_۲$ دو طرح نمونه‌گیری باشند و ثابت B وجود داشته باشد به‌طوری که به ازای تمام $\mathbf{x}$ها، $p_۱(\mathbf{x}) < Bp_۲(\mathbf{x})$ یک نمونه از $p_۲$ به‌عنوان نمونه‌ای از $p_۱$ پذیرفته می‌شود، اگر

$$U \leq \frac{p_۱(\mathbf{x})}{Bp_۲(\mathbf{x})}$$

که در آن $U \sim U(۰,۱)$، [۴۷].

تعریف ۱۵.۲ برآوردگر[۱۶] : یک برآوردگر، متغیری تصادفی است که اگر مقادیر نمونه‌ای در آن قرار داده شود، رخدادی از آماره متناظر با نمونه‌ی موجود به‌دست می آید. مقدار عددی مشاهده شده‌ی برآودگر را برآورد پارامتر گویند.

تعریف ۱۶.۲ برآوردگر مرتب[۱۷] : به برآوردگری که ترتیب واحدهای استخراج‌شده از جامعه را در نظر می‌گیرد، برآوردگر مرتب می‌گویند. بعضی از این برآوردگرها توسط داس[۱۸] [۱۴] در سال ۱۹۵۱ و دس راج[۱۹] [۴۴] در سال ۱۹۵٦ ارائه شده‌اند.

[۱۳] Bernoulli sampling design
[۱۴] Poisson sampling
[۱۵] Acceptance-Rejection technique
[۱٦] Estimator
[۱۷] Ordered estimator
[۱۸] Das, A.C.
[۱۹] Raj, D.

تعریف ۱۷.۲ برآوردگر نامرتب[۲۰] : متناظر با هر برآوردگر مرتب، برآوردگر نامرتبی نیز وجود دارد که به ترتیب استخراج واحدها بستگی ندارد. باسو[۲۱] و مورتی[۲۲] نشان دادند که چنین برآوردگری در مقایسه با برآوردگر مرتب کاراتر(دارای واریانس کمتر) خواهد بود. از جمله برآوردگرهای نامرتب، برآوردگر مورتی[۳۸] است.

تعریف ۱۸.۲ برآوردگر هارویتز–تامپسون (برآوردگر π)[۲۳] : هارویتز و تامپسون[۳۰] در سال ۱۹۵۲ با استفاده از احتمال شمول مرتبه‌ی اول، یک برآوردگر نااریب برای مجموع کل جامعه ارائه کردند که عبارت است از:

$$\hat{Y}_{HT} = \hat{Y}_{\pi} = \sum_{i \in s} \frac{y_i}{\pi_i}$$

که در آن π_i احتمال شمول مرتبه‌ی اول واحد i–اُم جامعه است. این برآوردگر در جوامع متناهی و خصوصاً در نمونه‌گیری بدون جایگذاری کاربرد دارد.

تعریف ۱۹.۲ برآوردگر هنسن–هرویتز (برآوردگر pwr)[۲۴] : اگر احتمال انتخاب عنصر k–اُم در هر مرحله از نمونه‌گیری p_k باشد و نمونه‌گیری با جایگذاری انجام شود، آن‌گاه فرآیند استخراج واحدها در هر مرحله مستقل از مراحل دیگر است. در چنین شرایطی هنسن و هرویتز[۲۶] در سال ۱۹۴۳ یک برآوردگر نااریب برای مجموع کل جامعه به شکل زیر معرفی کردند:

$$\hat{Y}_{HH} = \hat{Y}_{pwr} = \frac{1}{m} \sum_{i=1}^{m} \frac{y_{k_i}}{\mathrm{p}_{k_i}}$$

که در آن m اندازه‌ی نمونه‌ی استخراج‌شده و p_{k_i} احتمال انتخاب واحد k_i در نمونه است.

تعریف ۲۰.۲ آنتروپی[۲۵] : معروف‌ترین و پرکاربردترین آنتروپی، آنتروپی شانون[۵۳] است. فرض کنید متغیر تصادفی X، مقادیر $x_1, \ldots, x_n$ را با احتمال‌های $\mathrm{p}_1, \ldots, \mathrm{p}_n$ اختیار کند و $\sum_{i=1}^{n} \mathrm{p}_i = 1$. آنتروپی شانون متغیر تصادفی X به‌صورت $H(X) = -\sum_{i=1}^{n} \mathrm{p}_i(x) \log \mathrm{p}_i(x)$ تعریف می‌شود.

---
[۲۰] Unordered estimator

[۲۱] Basu, D.,

[۲۲] Murthy, M.N.

[۲۳] Horvitz-Thompson estimator(π estimator)

[۲۴] Hansen-Hurwitz estimator(pwr estimator)

[۲۵] Entropy

توسعه‌های گوناگونی روی آنتروپی توسط افراد مختلف صورت گرفته است که آنتروپی رنی[۲۶] [۴۶] معروف به آنتروپی مرتبه‌ی α،

$$H_\alpha(X) = \frac{۱}{۱-\alpha}\log(\sum_{i=۱}^{n} \mathrm{p}_i^\alpha)$$

و آنتروپی تی‌سالیس[۲۷] [۶۰]

$$H_\alpha(X) = \frac{۱}{۱-\alpha}(\sum_{i=۱}^{n} \mathrm{p}_i^\alpha - ۱)$$

از جمله معروف‌ترین آن‌ها است که در هر دو مورد $\lim_{\alpha\to۱} H_\alpha(X)$ آنتروپی شانون را نتیجه می‌دهد. همانند حالت گسسته، آنتروپی برای متغیر تصادفی پیوسته نیز تعریف شده است[۵۳].

تعریف **۲.۲۱ آنتروپی طرح نمونه‌گیری**[۲۸] : آنتروپی یک طرح نمونه‌گیری، اندازه‌ای از میزان تصادفی بودن طرح را نشان می‌دهد. آنتروپی طرح نمونه‌گیری p به‌صورت زیر تعریف می‌شود:

$$H = -\sum_{s\in\mathcal{Q}} p(s)\log p(s)$$

۲–۲ طرح‌های نمونه‌گیری

در این بخش به معرفی مختصری از طرح‌های نمونه‌گیری شناخته شده و ساختار جامعه‌ی متناظر آن‌ها پرداخته می‌شود. عمده‌ی مطالب این بخش مبتنی بر مراجع [۵۷] و [۶۲] است.
طرح‌های نمونه‌گیری را از لحاظ تعداد مراحل پیمایش می‌توان به دو رده‌ی طرح‌های یک‌مرحله‌ای و طرح‌های چندمرحله‌ای تفکیک نمود. همچنین از لحاظ تعداد انتخاب یک عنصر در نمونه، می‌توان دو رده‌ی طرح‌های با جایگذاری و بدون جایگذاری را در نظر گرفت. تقسیم‌بندی دیگر می‌تواند از لحاظ احتمال انتخاب عناصر باشد که در این صورت طرح‌های نمونه‌گیری با احتمال برابر و طرح‌های نمونه‌گیری با احتمال نابرابر را خواهیم داشت. از جمله طرح‌های نمونه‌گیری معروف عبارت‌اند از:

- **طرح‌های نمونه‌گیری ساده** $(simple)$[۲۹]

یک طرح نمونه‌گیری با پارامتر $\theta \in \mathbb{R}^+$ روی تکیه‌گاه $\mathcal{Q}$ ساده گفته می‌شود، اگر

---
[۲۶] Renyi, A.
[۲۷] Tsallis, C.
[۲۸] Entropy of sampling design
[۲۹] Simple sampling design