Impact Factor: 9.9 ISSN-L: 2544-980X

Об Одной Свойстве Статистики Хи- Квадрат

Абдуллаев Абдумухтор Ганиевич [1]

Аннотация: Хи-квадрат статистикаси, К. Пирсон томонидан киритилган бўлиб, гипотезаларни текшириш учун статистик таҳлилда кенг қўлланилади. Бироқ, унинг А. Реньи маъносинидаги аралашув (перемешивание) хусусияти, айниқса тасодифий миқдорлар кетма-кетлиги нуқтаи назаридан, етарлича ўрганилмаган. Ушбу тадқиқот айнан шу бўшлиқни тўлдиришга қаратилган бўлиб, хи-квадрат статистикаси учун R-аралашув хусусиятини қатъий математик ёндашув асосида исботлашни мақсад қилади. Методология мустақил кузатувлар кетма-кетлигини қуриш, уларнинг қийматларини кесишмас интервалларга ажратиш ва гуруҳланган частоталарнинг асимптотик тақсимотини таҳлил қилишни ўз ичига олади. Чегаравий эҳтимоллик усуллари, жумладан, Чебишев тенгсизлиги ва Крамер леммасидан фойдаланган ҳолда, тадқиқот хи-квадрат статистикасининг муайян шартлар бажарилганда R-аралашув хусусиятига эга эканлигини тасдиқлайди. Натижалар шуни кўрсатадики, танлама ҳажми ошган сари, хи-квадрат статистикасининг тақсимоти аралашув хусусиятини намоён қилади ва бу А. Реньи томонидан таклиф қилинган назарий асосларни тасдиқлайди. Олинган хулосалар статистик моделлаштириш ва гипотезаларни текширишда муҳим аҳамиятга эга бўлиб, айниқса, кузатувлар ўртасидаги боғлиқликлар муҳим роль ўйнаган ҳолларда жуда фойдали бўлиши мумкин. Тадқиқот статистик яқинлашиш (конвергенция) хусусиятларини чуқурроқ тушунишга ҳисса қўшади ва эҳтимоллар назарияси ҳамда амалий статистика соҳасидаги келгуси изланишлар учун янги имкониятлар яратади.

Калит сўзлар: Тасодифий миқдор, Пирсон –статистикаси, хи-квадрат критерийси, хи-квадрат таксимоти, А. Реньи, хосса, “силжиб юришлик” , тасодифий миқдорлар кетма-кетлиги.

ВВЕДЕНИЕ

Статистика хи-квадрат, введенная К. Пирсоном в 1990 году, широко используется в статистическом анализе, в частности для проверки соответствия распределений и выявления зависимости между категориальными переменными. На протяжении многих лет было проведено множество исследований, расширяющих работу Пирсона и изучающих различные статистические свойства данного критерия. Однако одним из недостаточно исследованных аспектов статистики хи-квадрат является её свойство перемешивания в смысле А. Реньи, имеющее важное значение для теории вероятностей и статистического вывода.

Несмотря на обширную литературу, посвященную распределению хи-квадрат и его приложениям, свойство статистического перемешивания, особенно в контексте последовательностей случайных величин, остается областью, требующей дальнейшего изучения. Настоящее исследование заполняет этот пробел, строго доказывая свойство перемешивания для статистики хи-квадрат в рамках подхода Реньи. Это вносит вклад в развитие теории статистических распределений и их свойств сходимости.

Для достижения этой цели в работе используется строгий математический подход, включающий анализ последовательностей независимых случайных величин, изучение свойств распределения хи-квадрат и принципов сходимости вероятностей. Исследование базируется на фундаментальных результатах Пирсона, расширяя их с применением определения перемешивания по Реньи и дополнительных теорем и лемм теории вероятностей.

[1] доцент негосударственного высшего учебного заведения “University of economics and pedagogy” (г. Андижан) abdullaev7591@gmail.com

Полученные результаты подтверждают, что при выполнении определенных условий последовательности независимых случайных величин удовлетворяют свойству R-перемешивания. Это имеет важные последствия для статистического моделирования, проверки гипотез и анализа зависимых наблюдений. Данное исследование углубляет теоретическое понимание статистики хи-квадрат и открывает новые перспективы для дальнейших исследований в области статистического перемешивания и его применения в современных статистических методах.

Методы

Методология данного исследования основана на строгом математическом подходе к доказательству свойства перемешивания статистики хи-квадрат в смысле А. Реньи. В работе применяется сочетание теоретического анализа, теории вероятностей и статистического вывода для расширения фундаментальных результатов, введенных К. Пирсоном. Исследование начинается с детального изучения последовательностей независимых случайных величин и их распределений хи-квадрат. Для построения теоретической базы используется формальное определение перемешивающихся последовательностей в интерпретации Реньи.

Доказательство строится поэтапно, начиная с рассмотрения последовательностей наблюдений, которые группируются в непересекающиеся интервалы. Асимптотическое распределение полученных частот анализируется в соответствии с распределением хи-квадрат с определенным числом степеней свободы. С использованием методов сходимости вероятностей и статистических неравенств, таких как неравенство Чебышева, исследование показывает, что последовательность случайных величин удовлетворяет свойству R-перемешивания. Для подтверждения основного результата применяются вспомогательные теоремы и леммы, включая результаты Крамера и другие фундаментальные работы. Выводы формируются на основе алгебраических преобразований и вероятностных аргументов, обеспечивающих необходимые условия сходимости.

Результаты и Обсуждение

Результаты исследования подтверждают, что при выполнении определенных условий статистика хи-квадрат обладает свойством перемешивания в смысле Реньи, что вносит вклад в более глубокое понимание свойств статистической сходимости. Такой методологический подход обеспечивает прочную теоретическую основу, позволяя расширить применение полученных результатов на более широкие статистические задачи, включая проверку гипотез и моделирование зависимых случайных величин в сложных статистических структурах. Данные результаты могут быть полезны для дальнейших исследований в области статистической теории и ее практических применений.

Настоящая работа посвящена доказательству свойства перемешивания смысле А. Реньи для статистики хи-квадрат предложенной К. Пирсоном.

Пусть ξ_1, ξ_2, …- последовательность независимых наблюдений над случайной величины (с.в.), ξ, $P(\xi < x) = F(x)$ и k – некоторое натуральное число. Разобъем область значений с.в. ξ на k непересекающихся интервалов точками

$$-\infty = a_0 < a_1 < \cdots < a_{k-1} < a_k = +\infty$$

и пусть $p_i = \Delta F(a_i) = F(a_i) - F(a_{i-1}) > 0, i = 1,2, \dots, k;$ $\nu_{in} -$ частоты, полученные после группировки выбоки ξ_1, ξ_2, …ξ_n по выше указанным интервалам. Тогда как показано Пирсоном, распределение последователь-ности с.в.

$$X_n^2 = \sum_{i=1}^{k} \frac{(\nu_{in} - np_i)^2}{np_i}$$

при $n \to \infty$ стремится к распределению $H_{k-1}(x)$, где $H_{k-1}(x)$ -распределение хи-квадрат с $(k - 1)$ степенями свободы,т.е.

$$H_{k-1}(x) = \frac{1}{2^{\frac{k-1}{2}}\Gamma(\frac{k-1}{2})}\int_0^x u^{\frac{k-3}{2}} e^{-\frac{u}{2}} du, (x > 0).$$

Прежде чем сформулировать теорему приведем определению свойства перемешивания введенный впервые в работе А. Реньи

Определение ([2]). Последовательность с.в. $\{\eta_n\}$ обладает свойством R–перемешивания с предельной функцией распределения (ф.р.) $F(x)$, если для любого события $A \epsilon \mathcal{F}$, $P(A) > 0$

$$\lim_{n\to\infty} P(\eta_n < x/A) = F(x)$$

в каждой точке x являющейся точкой непрерывности ф.р. $F(x)$.

Теорема. Последовательность с.в. $\{\mathrm{X}_n^2\}$ обладает свойством R -перемещивания с предельной ф.р. $H_{k-1}(x)$, т.е. для любого $A \epsilon \mathcal{F}$, $P(A) > 0$

$$\lim_{n\to\infty} P(\mathrm{X}_n^2 < x/A) = H_{k-1}(x) .$$

При $A = \Omega$ из теоремы 1 следует результат Пирсона.

При докозательстве теоремы мы восползуемся следующей леммой доказанной в работе **[2].**

Лемма 1. Пусть $\{\eta_n\}$ – некоторая последовательность с.в. Если в каждой точке непрерывности ф.р. $F(x)$

$$\lim_{n\to\infty} P(\eta_n < x) = F(x),$$

то $\{\eta_n\}$ является R– перемещивающейся с предельной ф.р. $F(x)$ тогда и только тогда, когда

$$\lim_{n\to\infty} P(\eta_n < x, \eta_{\kappa} < x) = F(x)P(\eta_k < x)$$

для $k = 1,2$ .... и в каждой точке непрерывности ф.р. $F(x)$.

Доказательство теоремы 1. На основании критерии R-перемешивае-мости (лемма 1) достаточно показать, что при каждом

m ($m = 1,2, \dots$)

$$\lim_{n\to\infty} P(\mathrm{X}_n^2 < x, \mathrm{X}_m^2 < x) = H_{k-1}(x)P(\mathrm{X}_m^2 < x) \quad (1)$$

Для любого m $(m < n)$ положим

$$w_{mn} = \sum_{i=1}^k \frac{S_{im}^2 + 2S_{im}\sum_{j=m+1}^n \xi_{ij}}{np_i}, \text{ где}$$

$S_{im} = \sum_{j=1}^m \xi_{ij}$, $\xi_{ij} = I(\xi_j \epsilon [\, a_{i-1}, a_i)) - p_i$, $I(A)$- индикатор события A.

В силу следующих тривиальных соотношения

$$P(\eta_1 + \eta_2 \ge c_1 + c_2) \le P(\eta_1 \ge c_1) + P(\eta_2 \ge c_2) \quad (2)$$

$$P(\eta_3 \cdot \eta_4 \ge c_3 c_4) \le P(\eta_3 \ge c_3) + P(\eta_4 \ge c_4) , \quad (3)$$

где $\eta_1, \eta_2, \eta_3 > 0, \eta_4 > 0$ -некоторые с.в., а c_1, c_2, c_3, c_4 - неслучайные величины, и из неравенство Чебышева для любого $\varepsilon > 0$ $(\varepsilon < 1/2)$ имеем

$$P(|w_{mn}| \ge \varepsilon) \le \sum_{i=1}^k \left[P\left(|S_{im}| \ge \left(\frac{n\varepsilon p_i}{2k}\right)^{1/2}\right) + P\left(|S_{im}| \ge \left(\frac{n\varepsilon p_i}{4k}\right)^{1/2} n^{-\varepsilon}\right) + +P\left(\left|\sum_{j=m+1}^n \xi_{ij}\right| \ge \left(\frac{n\varepsilon p_i}{4k}\right)^{1/2} n^{\varepsilon}\right)\right] \le \frac{2k(k-1)}{n^{2\varepsilon}\varepsilon}\left[\frac{m}{n^{1-2\varepsilon}} + \frac{2m}{n^{1-4\varepsilon}} + 2(1-\frac{m}{n})\right]$$

откуда $w_{mn} \xrightarrow{p} 0$ при $n \to \infty$.

Несложные вычисления показывают, что с.в. $X_n^2 - w_{mn}$ не зависит от с.в. X_m^2 при каждом m. Следовательно , для достаточно больших n верно неравенство

$$P(X_n^2 - w_{mn} < x - \varepsilon)P(X_m^2 < x) - \varepsilon \leq P(X_n^2 < x, X_m^2 < x) \leq$$

$$\leq P(X_n^2 - w_{mn} < x + \varepsilon)P(X_m^2 < x) + \varepsilon \ (4)$$

Здесь и всюду далее, для краткости говоря « для достаточно больших n …», понимаем под этим, что « для заданного $\varepsilon > 0$ существует такой номер $n(\varepsilon)$, что для всех $n > n(\varepsilon)$…» (причем $n(\varepsilon)$ не всегда одно и то же).

Применив к (4) лемму Крамера и в силу произвольности $\varepsilon > 0$ получаем (1).

Таким образом доказана теорема.

Литература

1. **G. Cramer**, *Mathematical Methods of Statistics*. Moscow, Russia: Mir, 1975.
2. **A. Renyi and P. Revers**, "On mixing sequences of random variables," *Acta Math. Acad. Sci. Hung.*, vol. 9, pp. 389-393, 1958.
3. **M. Fitzpatrick and M. I. Stewart**, "Asymptotics for Markov Chain Mixture Detection," *arXiv preprint*, arXiv:2111.12224, 2021. [Online]. Available: https://arxiv.org/abs/2111.12224.
4. **H. D. Chiang, Y. Matsushita, and T. Otsu**, "Multiway Empirical Likelihood," *arXiv preprint*, arXiv:2108.04852, 2021. [Online]. Available: https://arxiv.org/abs/2108.04852.
5. **G. Prevedello and K. R. Duffy**, "Discrete Convolution Statistic for Hypothesis Testing," *arXiv preprint*, arXiv:2008.13657, 2020. [Online]. Available: https://arxiv.org/abs/2008.13657.
6. **Y. Xu, X. Gao, and X. Wang**, "Nonparametric Clustering of Mixed Data Using Modified Chi-square Tests," *arXiv preprint*, arXiv:1301.6270, 2013. [Online]. Available: https://arxiv.org/abs/1301.6270.
7. **D. Sharpe**, "Your Chi-Square Test Is Statistically Significant: Now What?," *Practical Assessment, Research, and Evaluation*, vol. 20, no. 8, 2015. [Online]. Available: https://doi.org/10.7275/8d0s-c818.
8. **T. M. Franke, T. Ho, and C. A. Christie**, "The Chi-Square Test: Often Used and More Often Misinterpreted," *American Journal of Evaluation*, vol. 33, no. 3, pp. 448-458, 2012. [Online]. Available: https://doi.org/10.1177/1098214011426594.
9. **J. Lin et al.**, "Chi-Square Test for Imprecise Data in Consistency Table," *Frontiers in Applied Mathematics and Statistics*, 2023. [Online]. Available: https://doi.org/10.3389/fams.2023.1279638.
10. **S. Rana and R. Singhal**, "Effect Size for Chi-Square Test," *Real Statistics Using Excel*, 2020. [Online]. Available: https://real-statistics.com/chi-square-and-f-distributions/effect-size-chi-square/.
11. **M. Bäckström**, "Fast Randomization for Distributed Low-Bitrate Coding of Speech and Audio," *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 26, no. 1, pp. 186-197, 2018. [Online]. Available: https://doi.org/10.1109/TASLP.2017.2761544.
12. **J. Bausch**, "On the Efficient Calculation of a Linear Combination of Chi-Square Random Variables with an Application in Counting String Vacua," *Journal of Physics A: Mathematical and Theoretical*, vol. 45, no. 20, 2012. [Online]. Available: https://doi.org/10.1088/1751-8113/45/20/205205.
13. **T. B. Bäckström**, "Fast Randomization for Distributed Low-Bitrate Coding of Speech and Audio," *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 26, no. 1, pp. 186-197, 2018. [Online]. Available: https://doi.org/10.1109/TASLP.2017.2761544.

14. **J. Bausch**, "On the Efficient Calculation of a Linear Combination of Chi-Square Random Variables with an Application in Counting String Vacua," *Journal of Physics A: Mathematical and Theoretical*, vol. 45, no. 20, 2012. [Online]. Available: https://doi.org/10.1088/1751-8113/45/20/205205.
15. **T. B. Bäckström**, "Fast Randomization for Distributed Low-Bitrate Coding of Speech and Audio," *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 26, no. 1, pp. 186-197, 2018. [Online]. Available: https://doi.org/10.1109/TASLP.2017.2761544.