|  |  |  |
| --- | --- | --- |
|  | | |
| Федеральное государственное бюджетное  образовательное учреждение высшего образования «Новосибирский государственный технический университет» | | |
|  | | |
| Кафедра прикладной математики | | |
| Практическое задание № 3 | | |
| по дисциплине «Методы принятия оптимальных решений» | | |
| **Идентификация законов распределения** | | |
|  | | |
|  | Бригада | ПМ-13 Буданцев дмитрий |
| . | ПМ-13 Форкин Кирилл |
|  |  |
|  |  |
|  |  |
|  |  |
| Преподаватели | Лемешко Борис Юрьевич |
|  |  |
| Новосибирск,2024 | | |

**Проведённые исследования**

1. **Подобрать закон распределения, наилучшим образом описывающий выборку K1000.dat.**
   1. **Построить по выборке модели симметричных законов (нормальный, логистический, двусторонний экспоненциальный), оценить для построенных законов те же вероятности, зафиксировать значения в построенной таблице. Выполнить то же, используя режим «группировать при оценивании».**

Результаты проверки **сложных гипотез** *симметричных* *моделей распределений**(см. таблица 1)* на основании предложенного файла с объемом выборки в **1000** наблюдений, опираясь на критерии согласия *Колмагорова*, *Крамера-Мизиса-Смирнова*, *Андесона-Дарлинга* и *Хи-квадрат Пирсона*.

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | Колмогоров | | КМС | | АД | | Хи-квадарат | |
| Модель  распределения | S | P | S | P | S | P | S | P |
| Нормальное | 7.46068 | 0 | 24.33152 | 0 | 2.00E+100 | 0 | 549821.9 | 0 |
| Логическое | 3.17958 | 0 | 3.8549 | 0 | 22.41637 | 0 | 58181.79 | 0 |
| Двух. Эксп. | 1.1361 | 0 | 0.30532 | 0 | 2.00226 | 0 | 441.8278 | 0 |

*Таблица 1.* Результаты проверки сложных гипотез выборки К1000 для симметричных законов

Из результатов (*см. таблица 1*) видно, что при проверке сложных гипотез критерии Колмогорова, Крамера-Мизеса-Смирнова, Андерсона-Дарлинга и Хи-квадрат Пирсона, не дают никаких оснований для принятия ложной гипотезы .

* 1. **Используя 2 режима построить несимметричные модели (минимального значения, Su-Джонсона) и оценить соответствующие вероятности.**

Результаты проверки **сложных гипотез** *несимметричных моделей распределений (см. таблица 2)* на основании предложенного файла с объемом выборки в **1000** наблюдений, опираясь на критерии согласия *Колмагорова*, *Крамера-Мизиса-Смирнова*, *Андесона-Дарлинга* и *Хи-квадрат Пирсона(см. таблица 3)*.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | Колмогоров | | КМС | | АД | |
| Модель распределения | S | P | S | P | S | P |
| Su-Джонсона | 0.52330 | 0.48441 | 0.03174 | 0.37049 | 0.27641 | 0.20673 |
| Мин. Значение | 9.35564 | 0 | 24.9189 | 0 | 2.00E+100 | 0 |

*Таблица 2.* Результаты проверки сложных гипотез выборки К1000 для непараметрических критериев

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| К1000 | Хи-квадарат (группирование при оценивании) | | Хи-квадарат (без группирование при оценивании) | |
| Модель распределения | S | P | S | P |
| Su-Джонсона | 10,4332 | 0,40334 | 13,976 | 0,234324 |
| Мин. Значение | 174,1053 | 6,90E-31 | 60663,85 | 0,00E+00 |

*Таблица 3.* Результаты проверки сложных гипотез выборки К1000

для критерия Хи-квадрат

Из результатов (*см. таблица 2-3*) видно, что при проверке сложных гипотез критерии Колмогорова, Крамера-Мизеса-Смирнова, Андерсона-Дарлинга и Хи-квадрат Пирсона, не отвергают гипотезу Su-Джонсона. При этом критерии отвергают гипотезум минимального значения. Можно заметить, что оценённые параметры ассиметрии модели Su-Джонсона близки к 0, что говорит о том, что полученное распределение близка к симметричной, что подтверждается графиком плотности *(см. график 1)*. Удивительно, что ассиметричная модель оказалась ближе, чем симметричные, но при таких параметрах она является почти симметричным.



*График 1.* График плотности Su-Джонсона и выборки К1000

* 1. Опираясь на непараметрические критерии, постарайтесь подобрать модель, наилучшим образом описывающую данную выборку.

По результату проверки сложных гипотез при различных моделях распределений было выявлена самая подходящая модель из тех, которые предложены в ISW *(см. таблица 3)*

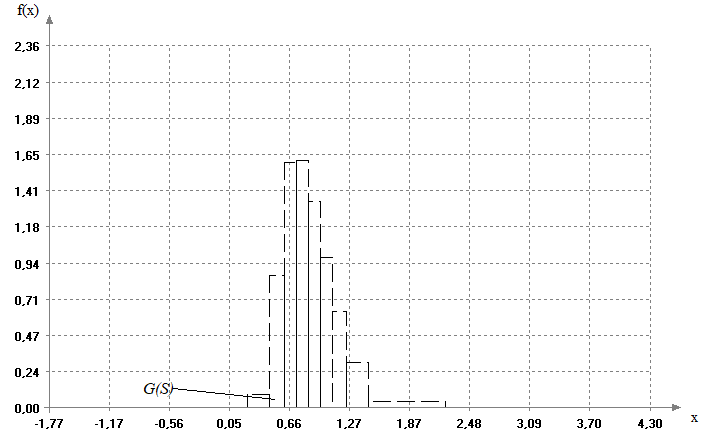
|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Модель распределения | Колмогоров | | КМС | | АД | | Хи-квадарат | |
| S | P | S | P | S | P | S | P |
| Коши | 0.36410 | 0.98564 | 0.01631 | 0.98422 | 0.12437 | 0.99777 | 5.04613 | 0.923 |

*Таблица 3.* Результаты проверки сложных гипотез выборки К1000 для распределения Коши

Таким образом, из *таблицы 3,* можно сделать вывод, что вероятнее всего предложенная выборка описывает модель распределения Коши. Как известно распределение Коши является симметричным законом, что согласуется с плотностью выборки, которая была представленна ранее *(см. график 1).* Данный пример также иллюстрирует то, что непараметрические критерии согласия, одинаково хорошо определяют модель распределения.

1. **Для выборки *model\_1.dat* подобрать наилучшим образом описывающую её параметрическую модель.**

В первую очередь обратим внимание на плотность выборки *(см. график 2*). Оно имеет скос в левую сторону, т.е. имеют асимметрию. Таким образом, подбирать симметричные модели не имеет, смыла.



*График 2.* График плотности выборки model\_1

Результаты попыток подобрать наилучшую параметрическую модель *(см. таблицу 4)*, для выборки *model\_1* с обёмом в **10000** наблюдений, опираясь на критерии согласия *Колмагорова*, *Крамера-Мизиса-Смирнова*, *Андесона-Дарлинга* и *Хи-квадрат Пирсона*.

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Модель  распределения | Колмогоров | | КМС | | АД | | Хи-квадарат | |
| S | P | S | P | S | P | S | P |
| Гамма | 0.67179 | 0 | 0.082867 | 0 | 0.51862 | 0 | 13.38301 | 0.781 |
| Бета 2 рода | 1.43789 | 0 | 0.37159 | 0 | 2.74912 | 0 | 48.40324 | 0 |
| Бета 3 рода | 0.68364 | 0 | 0.06204 | 0 | 0.39011 | 0 | 12.45154 | 0.496 |
| Su-Джонсона | 2.46148 | 0 | 2.33445 | 0 | 17.75676 | 0 | 13.38301 | 0.782 |
| Sl-Джонсона | 1.56798 | 0 | 0.48203 | 0 | 3.19428 | 0 | 35.49395 | 0.0002 |

*Таблица 4.* Результаты проверки сложных гипотез выборки model\_1

Из результатов (*см. таблица 4*) видно, что при проверке сложных гипотез критерии Колмогорова, Крамера-Мизеса-Смирнова, Андерсона-Дарлинга, отвергают проверяемые гипотезы, но критерий Хи-квадрат не отвергает большинство гипотез. Так, как непараметрические критерии имеют большую мощность при проверки сложных гипотез ясно, что данные модели не описывают данную выборку.

До того пока я узнал, что данная выборка соответствуют распределению Колмогорова, я не смог подобрать истинную модель для данной выборки. Причина в том, что я не загрузил эту модель и не пробовал проверять гипотезу с распределением Колмогорова. В итоге я лишь нашёл только самые похожие на данное распределение.

1. **Подобрать параметрическую модель, наилучшим образом описывающую выборку model\_2.dat. Аналогично п.2.**

Действуя аналогично пункту 2. Я сделал вывод, что для данной выборки лучше всего подбирать несимметричные модели.

Результаты попыток подобрать наилучшую параметрическую модель *(см. таблицу 5)*, для выборки *model\_2* с обёмом в **10000** наблюдений, опираясь на критерии согласия *Колмагорова*, *Крамера-Мизиса-Смирнова*, *Андесона-Дарлинга* и *Хи-квадрат Пирсона*.

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Модель  распределения | Колмогоров | | КМС | | АД | | Хи-квадарат | |
| S | P | S | P | S | P | S | P |
| Бета 3 рода | 0.99627 | 0 | 0.22916 | 0 | 1.40208 | 0 | 14.80903 | 0.322 |
| Sl-Джонсона | 1.93136 | 0 | 1.0269 | 0 | 6.2015 | 0 | 18.21734 | 0.449 |

*Таблица 5.* Результаты проверки сложных гипотез выборки model\_2

Анологично 2 пункту, из результатов (*см. таблица 4*) видно, что при проверке сложных гипотез критерии Колмогорова, Крамера-Мизеса-Смирнова, Андерсона-Дарлинга, отвергают проверяемые гипотезы, но критерий Хи-квадрат не отвергает гипотезы. Так, как непараметрические критерии имеют большую мощность при проверки сложных гипотез ясно, что данные модели не описывают данную выборку.

Анологично 2 пункту до того пока я узнал, что данная выборка соответствуют распределению a1, я не смог подобрать истинную модель для данной выборки. Причина в том, что я не загрузил эту модель и не пробовал проверять гипотезу с распределением a1. В итоге я лишь нашёл только самые похожие на данное распределение.

1. **Подобрать параметрическую модель, наилучшим образом описывающую выборку model\_3.dat. Аналогично п.2.**

Действуя аналогично пункту 2. Я сделал вывод, что для данной выборки лучше всего подбирать несимметричные модели.

Результаты попыток подобрать наилучшую параметрическую модель *(см. таблицу 6)*, для выборки *model\_3* с обёмом в **10000** наблюдений, опираясь на критерии согласия *Колмагорова*, *Крамера-Мизиса-Смирнова*, *Андесона-Дарлинга* и *Хи-квадрат Пирсона*.

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Модель  распределения | Колмогоров | | КМС | | АД | | Хи-квадарат | |
| S | P | S | P | S | P | S | P |
| Бета 3 рода | 0.77001 | 0 | 0.08867 | 0 | 0.56768 | 0 | 12.45154 | 0.496 |
| Sl-Джонсона | 0.67029 | 0 | 0.07032 | 0 | 0.593 | 0 | 12.45194 | 0.386 |

*Таблица 6.* Результаты проверки сложных гипотез выборки model\_3

Анологично 2, из результатов (*см. таблица 6*) видно, что при проверке сложных гипотез критерии Колмогорова, Крамера-Мизеса-Смирнова, Андерсона-Дарлинга, отвергают проверяемые гипотезы, но критерий Хи-квадрат не отвергает гипотезы. Так, как непараметрические критерии имеют большую мощность при проверки сложных гипотез ясно, что данные модели не описывают данную выборку.

Анологично 2 пункту до того пока я узнал, что данная выборка соответствуют распределению a2, я не смог подобрать истинную модель для данной выборки. Причина в том, что я не загрузил эту модель и не пробовал проверять гипотезу с распределением a2. В итоге я лишь нашёл только самые похожие на данное распределение.

1. **Подобрать параметрическую модель, наилучшим образом описывающую выборку model\_4.dat. Аналогично п.2.**

Действуя аналогично пункту 2. Я сделал вывод, что для данной выборки лучше всего подбирать несимметричные модели.

Результаты попыток подобрать наилучшую параметрическую модель *(см. таблицу 7)*, для выборки *model\_4* с обёмом в **10000** наблюдений, опираясь на критерии согласия *Колмагорова*, *Крамера-Мизиса-Смирнова*, *Андесона-Дарлинга* и *Хи-квадрат Пирсона*.

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Модель распределения | Колмогоров | | КМС | | АД | | Хи-квадарат | |
| S | P | S | P | S | P | S | P |
| Su-Джонсона | 1.71632 | 0 | 0.66599 | 0 | 4.22235 | 0 | 16.91233 | 0.338 |
| Бета 2 рода | 0.46730 | 0 | 0.02101 | 0 | 0.15829 | 0 | 11.35225 | 0.46 |
| Бета 3 рода | 0.83853 | 0 | 0.16421 | 0 | 1.16674 | 0 | 14.08674 | 0.286 |

*Таблица 7.* Результаты проверки сложных гипотез выборки model\_4

Анологично 2, из результатов (*см. таблица 7*) видно, что при проверке сложных гипотез критерии Колмогорова, Крамера-Мизеса-Смирнова, Андерсона-Дарлинга, отвергают проверяемые гипотезы, но критерий Хи-квадрат не отвергает гипотезы. Так, как непараметрические критерии имеют большую мощность при проверки сложных гипотез ясно, что данные модели не описывают данную выборку.

Анологично 2 пункту до того пока я узнал, что данная выборка соответствуют распределению Фишера, я не смог подобрать истинную модель для данной выборки. Причина в том, что я не загрузил эту модель и не пробовал проверять гипотезу с распределением Фишера. В итоге я лишь нашёл только самые похожие на данное распределение.

Хоть в пунктах 2-4, не удалось найти истинной модели, но похожие модели подобрать удалось.

1. **Проверьте простые гипотезы о принадлежности выборок «истинным» законам, в соответствии с которыми генерировались соответствующие предложенные выборки.**

Так как в *1 пункте* модель подобрана верно, проверять гипотезу ещё раз не вижу смысла. Распределение Фишера нет в ISW, поэтому проверки гипотезы не будут приведены.

Результаты проверки **простых гипотез** *истенных моделей распределений (см. таблица 8)* на основании предложенных файлов с объемом выборки в **10000** наблюдений, опираясь на критерии согласия *Колмагорова*, *Крамера-Мизиса-Смирнова*, *Андесона-Дарлинга* и *Хи-квадрат Пирсона*.

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Модель распределения и имя выборки | Колмогоров | | КМС | | АД | | Хи-квадарат | |
| S | P | S | P | S | P | S | P |
| Колмогорова (model\_1) | 0.63580 | 0.81362 | 0.0593 | 0.81836 | 0.56499 | 0.68168 | 13.57751 | 0.482 |
| a1 (model\_2) | 0.63580 | 0.81362 | 0.0593 | 0.81836 | 0.56499 | 0.68168 | 12.551 | 0.56214 |
| a2 (model\_3) | 0.63580 | 0.81362 | 0.0593 | 0.81836 | 0.56499 | 0.68168 | 13.57751 | 0.482 |

*Таблица 8.* Результаты проверки простых гипотез предложенных выборок на основе истинных моделей

Из *таблицы 8*, видно, как отличается мощность непараметрических критериев с Хи-квадрат. Непараметрические критерии по итогам исследования приняли только истинную гипотезу, в отличае от Хи-квадрат, что в очередной раз показывает, что непараметрические критерии не зависят от размера выборки и становятся только мощней от её увеличения.

1. **«Считая», что наилучшая построенная Вами модель получена по некоторой другой выборке, проверьте согласие с Вашими моделями. Сравните достигаемые уровни значимости, с полученными в п. 6.**

Буду считать, что наилучшая модель будет характеризоваться, большим уровнем согласия, полученным у критерия Хи-квадрат.

Результаты проверки **простых гипотез “**наилучших моделей” *(см. таблица )*, полученные в результате исследования, на основании предложенных выборок с объёмом **10000** наблюдений, опираясь на критерии согласия *Колмагорова*, *Крамера-Мизиса-Смирнова*, *Андесона-Дарлинга* и *Хи-квадрат Пирсона*.

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Модель распределения и имя выборки | Колмогорова | | КМС | | АД | | Хи-квадрат | |
| S | P | S | P | S | P | S | P |
| Гамма (4.7475, 0.1210, 0.2948) (model\_1) | 0,65256 | 0,78803 | 0,081713 | 0,682003 | 0,520793 | 0,72569 | 10,394 | 0,732847 |
| Sl-Джонсона (0.1623, 1.1289, 0.1302, 0.0085) (model\_2) | 1,264813 | 0,081559 | 0,403925 | 0,070545 | 3,321631 | 0,018845 | 21,7167 | 0,084626 |
| Бета 3-го рода (3.7351, 2.5510, 16.6547, 8.2133, 0.0894) (model\_3) | 0,768558 | 0,596035 | 0,088375 | 0,644899 | 0,566536 | 0,680179 | 14,02526 | 0,4478 |
| Бета 2-го рода (2.4722, 9.7451, 3.9789, -0.0043) (model\_4) | 0,485186 | 0,972638 | 0,032691 | 0,966514 | 0,234494 | 0,977978 | 11,34954 | 0,6584 |

*Таблица 9.* Результаты проверки простых гипотез предложенных выборок на основе “наилучших моделей”

Не сложно заметить, что в сравнении с “наилучшими моделями”, которые были получены в процессе исследования, непараметрические критерии согласия имеют не нулевое значение и не отвергают гипотезу. Видно, что значения Хи-квадрат, продолжает не отвергать гипотезу. В целом как было выяснено в прошлых работах, критерий Хи-квадрат имеет меньшую мощность, чем непараметрические критерии при проверке сложных гипотез. И ситуация, которая была получена во *2-4 пункте*, этим объясняется. При проверке простых гипотез с методом группировки АОГ, мощность критерия Хи-квадрат больше некоторых непараметрических критериев. И результат, полученный в *таблице 8,* так же объясним.