

МИНИСТЕРСТВО НАУКИ И ВЫСШЕГО ОБРАЗОВАНИЯ РОССИЙСКОЙ
ФЕДЕРАЦИИ
ФЕДЕРАЛЬНОЕ ГОСУДАРСТВЕННОЕ БЮДЖЕТНОЕ
ОБРАЗОВАТЕЛЬНОЕ УЧРЕЖДЕНИЕ
ВЫСШЕГО ОБРАЗОВАНИЯ
«НОВОСИБИРСКИЙ ГОСУДАРСТВЕННЫЙ ТЕХНИЧЕСКИЙ
УНИВЕРСИТЕТ»

Кафедра систем сбора и обработки данных

ЛАБОРАТОРНАЯ РАБОТА №5

по дисциплине: Компьютерные технологии моделирования и анализа данных
на тему: Экспериментальное исследование предельных распределений
статистик непараметрических критериев согласия. Часть 1.

Вариант №2

Факультет: ФПМИ

Группа: ПММ-21

Выполнил: Сухих А.С., Черненко Д.А.

Проверил: д.т.н., профессор Лемешко Борис Юрьевич

Дата выполнения: 08.12.22

Отметка о защите:

Новосибирск 2022

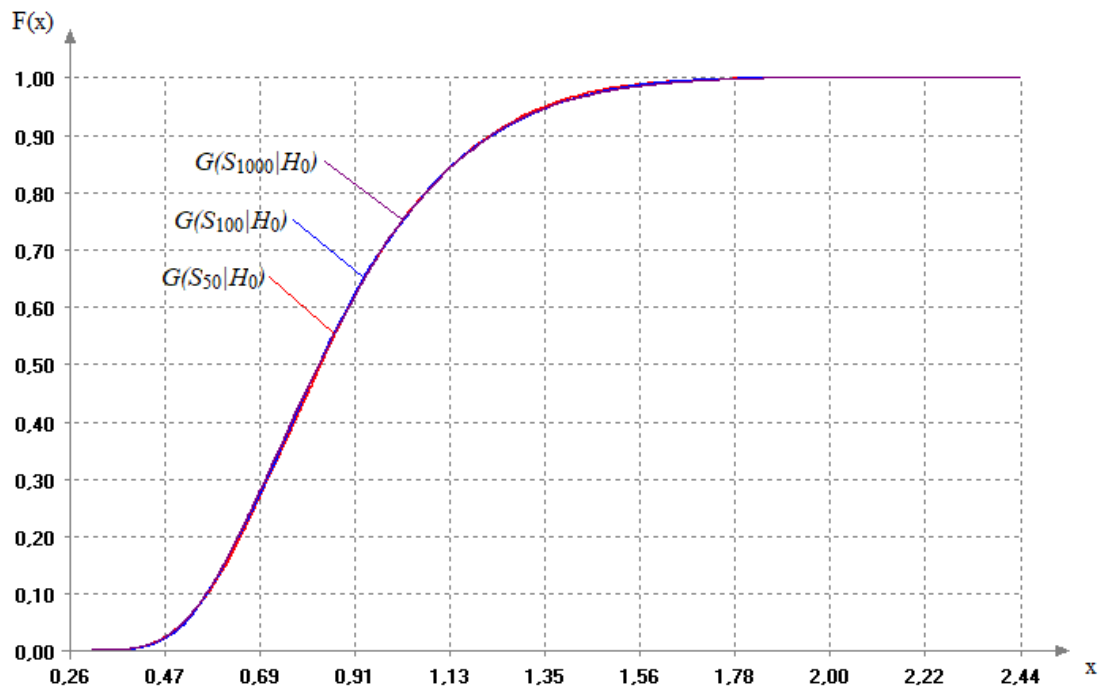
Цель работы: Исследование распределений статистик непараметрических критериев согласия при проверке простых и различных сложных гипотез. В первой части исследуются распределения статистик критериев согласия Колмогорова, Смирнова, ω^2 Крамера–Мизеса–Смирнова и \mathcal{Q}^2 Андерсона–Дарлинга.

Ход работы:

1. Смоделировать распределение статистики S для заданного критерия согласия при простой гипотезе H_0 . Сравнить полученное эмпирическое распределение с предельным распределением классической статистики.

Колмогоров:

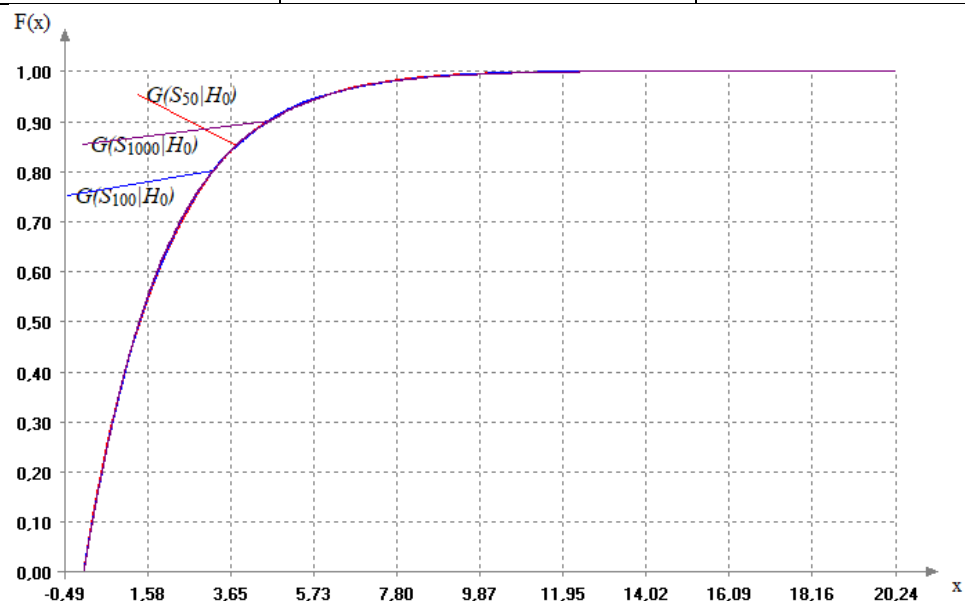
$n = 50$	$n = 100$	$n = 1000$
$P = 1 - G(S H_0)$ $= 0.006433097934900212$ ОТВЕРГАЕТСЯ	$P = 1 - G(S H_0)$ $= 0.2625263117932166$ ОТВЕРГАЕТСЯ	$P = 1 - G(S H_0)$ $= 0.7094836420769316$ НЕ ОТВЕРГАЕТСЯ



● Критерий Колмогорова $N=16600$ $G(S|H_0)$ H_0 Лог(0.0000,1.0000) $n=50$ ОМР ГСЧ=100
 ● Критерий Колмогорова $N=16600$ $G(S|H_0)$ H_0 Лог(0.0000,1.0000) $n=100$ ОМР ГСЧ=100
 ● Критерий Колмогорова $N=16600$ $G(S|H_0)$ H_0 Лог(0.0000,1.0000) $n=1000$ ОМР ГСЧ=100

Смирнов:

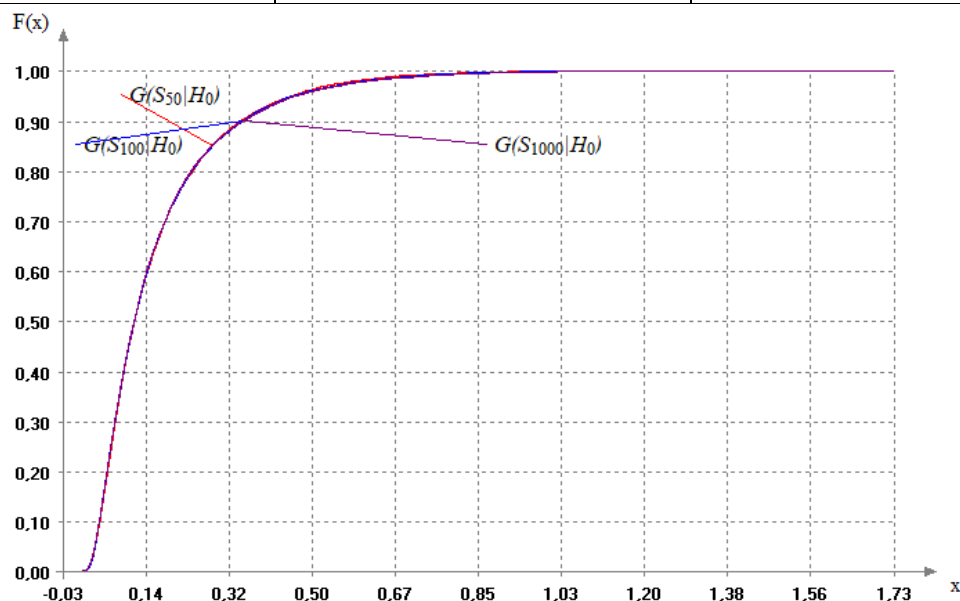
$n = 50$	$n = 100$	$n = 1000$
$P = 1 - G(S H_0)$ $= 0.1149030644691371$ НЕ ОТВЕРГАЕТСЯ	$P = 1 - G(S H_0)$ $= 0.4053445270498439$ НЕ ОТВЕРГАЕТСЯ	$P = 1 - G(S H_0)$ $= 0.8287675049284802$ НЕ ОТВЕРГАЕТСЯ



● Критерий Смирнова N=16600 G(SIH0) H0 Лог(0.0000,1.0000) n=50 OMP ГСЧ=100
 ● Критерий Смирнова N=16600 G(SIH0) H0 Лог(0.0000,1.0000) n=100 OMP ГСЧ=100
 ● Критерий Смирнова N=16600 G(SIH0) H0 Лог(0.0000,1.0000) n=1000 OMP ГСЧ=100

Крамер-Мизес-Смирнов:

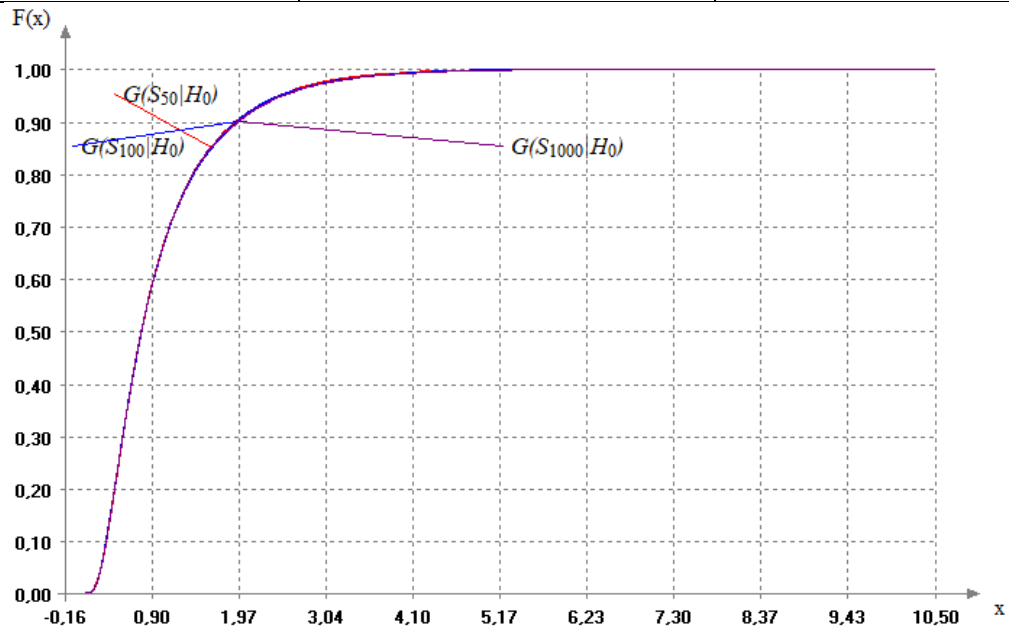
$n = 50$	$n = 100$	$n = 1000$
$P = 1 - G(S H_0)$ $= 0.01879968988182557$ ОТВЕРГАЕТСЯ	$P = 1 - G(S H_0)$ $= 0.2316364700061102$ НЕ ОТВЕРГАЕТСЯ	$P = 1 - G(S H_0)$ $= 0.560153762517235$ НЕ ОТВЕРГАЕТСЯ



● Критерий Крамера-Мизеса-Смирнова N=16600 G(SIH0) H0 Лог(0.0000,1.0000) n=50 OMP ГСЧ=100
 ● Критерий Крамера-Мизеса-Смирнова N=16600 G(SIH0) H0 Лог(0.0000,1.0000) n=100 OMP ГСЧ=100
 ● Критерий Крамера-Мизеса-Смирнова N=16600 G(SIH0) H0 Лог(0.0000,1.0000) n=1000 OMP ГСЧ=100

Андерсон-Дарлинг:

$n = 50$	$n = 100$	$n = 1000$
$P = 1 - G(S H_0)$ $= 0.2466149878141375$ НЕ ОТВЕРГАЕТСЯ	$P = 1 - G(S H_0)$ $= 0.3370570681615366$ НЕ ОТВЕРГАЕТСЯ	$P = 1 - G(S H_0)$ $= 0.3768096827904253$ НЕ ОТВЕРГАЕТСЯ



● Критерий Андерсона-Дарлинга N=16600 $G(S|H_0) \sim \text{Log}(0.0000, 1.0000)$ n=50 OMP ГСЧ=100
● Критерий Андерсона-Дарлинга N=16600 $G(S|H_0) \sim \text{Log}(0.0000, 1.0000)$ n=100 OMP ГСЧ=100
● Критерий Андерсона-Дарлинга N=16600 $G(S|H_0) \sim \text{Log}(0.0000, 1.0000)$ n=1000 OMP ГСЧ=100

2. Смоделировать распределение статистики S для этого же критерия согласия при сложной гипотезе H_0 . Попытаться идентифицировать полученное эмпирическое распределение, используя систему статистического анализа ISW.

Колмогоров:

$n = 50$	$n = 100$	$n = 1000$
Хи-Квадрат Пирсона: Рав (0.2874, 0.8780)	Хи-Квадрат Пирсона (критерий согласия): Гамма (6.8339, 0.0462, 0.2645)	Хи-Квадрат Пирсона (критерий согласия): Гамма (7.4822, 0.0456, 0.2410)
Колмогоров: Рав(0.0682, 1.4759)	Колмогоров: Равн (0.2662, 0.9572)	Колмогоров: Логарифмическое (ln) N(-0.5626, 0.2069)

Смирнов:

$n = 50$	$n = 100$	$n = 1000$
Хи-Квадрат Пирсона: Би-П (6.6541, 8.4869, 1.1414, 0.1730) Смирнов: а1	Хи-Квадрат Пирсона: Бе-П (5.8981, 6.8076, 0.9667, 0.2188) Смирнов: а1	Хи-Квадрат Пирсона: SI-Джонсон (0.2672, 2.0058, 1.0982, 0.1047) Смирнов: Мин(0.0000, 1.0000)

Крамер-Мизес-Смирнов:

$n = 50$	$n = 100$	$n = 1000$
Хи-Квадрат Пирсона: Гамма (2.4632, 0.0161, 0.0085) К-М-С: Рав (0.0085, 0.2419)	Хи-Квадрат Пирсона: Гамма (2.4288, 0.0160, 0.0093) К-М-С: Рав (0.0093, 0.2668)	Хи-Квадрат Пирсона: Рав (0.0068, 0.2415) К-М-С: Рав (0.0068, 0.2415)

Андерсон-Дарлинг:

$n = 50$	$n = 100$	$n = 1000$
Хи-Квадрат Пирсона: Логарифмический (ln) N(-1.1509, 0.4314) Анд-Дар: Рав(0.0682, 1.4759)	Хи-Квадрат Пирсона: Логарифмический (ln) N(-1.1436, 0.4282) Анд-Дар: Рав (0.0766, 1.6648)	Хи-Квадрат Пирсона: Логарифмический (ln) N(-1.1393, 0.4331) Анд.-Дарл: Рав (0.0637, 1.4969)

Получилась достаточно интересная ситуация: помимо распределения, основанного на критерии согласия Колмогорова с $n = 1000$, ни одно другое эмпирическое распределение не удалось наилучшим образом аппроксимировать с наиболее подходящими распределениями, в предельных случаях.

Можно предположить, что для случаев с $n = 50, 100$ этого не удалось достичь ввиду небольшого объёма выборки, из-за чего не удалось получить предельные распределения.

3. Смоделировать распределения статистики S исследуемого критерия согласия при проверке простой и сложной гипотез H_0 при справедливой гипотезе H_1 . Для того чтобы распределение,

соответствующее гипотезе H_1 было наиболее близким к распределению, соответствующему гипотезе H_0 , следует подобрать параметры распределения, соответствующего гипотезе H_1 , из условия минимизации расстояния до распределения, соответствующего основной гипотезе.

4. Построить оперативные характеристики критерия для простой и сложных гипотез как функции вида $(1-\beta)(\alpha)$. Сравнить мощность непараметрических критериев с мощностью критериев χ^2 Пирсона и Рао-Робсона-Никулина (оперативные характеристики критериев χ^2 Пирсона и Рао-Робсона-Никулина взять из отчетов по лабораторным работам № 3-4).

Для Пирсона АОГ, $k = 10$:

α	Простая гипотеза			Сложная гипотеза		
	$1 - \beta$					
	$n = 50$	$n = 100$	$n = 1000$	$n = 50$	$n = 100$	$n = 1000$
0.15	0.199639	0.248614	0.902229	0,153373	0,207892	0,896024
0.1	0.139699	0.17994	0.857169	0,101988	0,141566	0,845361
0.05	0.073253	0.101386	0.772831	0,0512048	0,0759036	0,743133
0.025	0.0372892	0.0570482	0.685241	0,0246988	0,0392169	0,63241
0.01	0.0155422	0.025	0.560964	0,0109036	0,0157831	0,476928

Для Никулина АОГ, $k = 10$:

α	Простая гипотеза			Сложная гипотеза		
	$1 - \beta$					
	$n = 50$	$n = 100$	$n = 1000$	$n = 50$	$n = 100$	$n = 1000$
0.15	0.199699	0.248494	0.902169	0.148373	0.190361	0.87512
0.1	0.139639	0.17988	0.857048	0.0968675	0.128494	0.818072
0.05	0.0733133	0.101265	0.772711	0.0453012	0.0672892	0.707952
0.025	0.0372289	0.056988	0.68506	0.0213855	0.0328916	0.585
0.01	0.0155422	0.0249398	0.560663	0.00777108	0.0121687	0.439157

Для Колмогорова:

α	Простая гипотеза			Сложная гипотеза		
	$1 - \beta$					
	$n = 50$	$n = 100$	$n = 1000$	$n = 50$	$n = 100$	$n = 1000$
0.15	0.19012	0.221386	0.666145	0.191747	0.236265	0.828253
0.1	0.129819	0.150361	0.554337	0.133373	0.164398	0.753614
0.05	0.0712048	0.0819277	0.373072	0.0687349	0.0930723	0.609217
0.025	0.0418675	0.0461446	0.238193	0.0348795	0.0516265	0.476506
0.01	0.0196386	0.0224096	0.124518	0.0145181	0.0242169	0.311205

Для Смирнова:

α	Простая гипотеза			Сложная гипотеза		
	$1 - \beta$					
	$n = 50$	$n = 100$	$n = 1000$	$n = 50$	$n = 100$	$n = 1000$
0.15	0.177349	0.203855	0.541084	0.181807	0.235964	0.784277
0.1	0.124398	0.146145	0.434819	0.123976	0.169337	0.700422
0.05	0.0661446	0.0792771	0.298253	0.0681928	0.0904819	0.556024
0.025	0.035241	0.0416265	0.199217	0.0363855	0.0495181	0.426867
0.01	0.0187952	0.0198795	0.107048	0.015	0.0211446	0.267831

Для Крамера-Мизеса-Смирнова:

α	Простая гипотеза			Сложная гипотеза		
	$1 - \beta$					
	$n = 50$	$n = 100$	$n = 1000$	$n = 50$	$n = 100$	$n = 1000$
0.15	0.18012	0.203614	0.682289	0.197349	0.254337	0.918976
0.1	0.125482	0.136566	0.533795	0.133614	0.184819	0.872711
0.05	0.0671084	0.0746988	0.329337	0.0693373	0.103795	0.77753
0.025	0.0378916	0.0399398	0.183253	0.035	0.0613855	0.660904
0.01	0.018494	0.0161446	0.0728313	0.016506	0.0254217	0.516867

Для Андерсона-Дарлинга:

α	Простая гипотеза			Сложная гипотеза		
	$1 - \beta$					
	$n = 50$	$n = 100$	$n = 1000$	$n = 50$	$n = 100$	$n = 1000$
0.15	0.187952	0.21512	0.749157	0.178795	0.240482	0.954458
0.1	0.132289	0.14759	0.626024	0.11988	0.17259	0.92247
0.05	0.0683735	0.0820482	0.427711	0.0624699	0.0975301	0.846446
0.025	0.0390361	0.0425301	0.271627	0.0316265	0.0521084	0.75241
0.01	0.0186145	0.0174096	0.129699	0.0121084	0.0212651	0.614398

Вывод:

При простой гипотезе с $n = 100, 1000$ непараметрические критерии согласия работают лучше параметрических, что касается ситуации с простыми гипотезами с $n = 50$, можно сказать, что параметрические и непараметрические критерии согласия работают примерно одинаково.

При сложных гипотезах с $n = 50, 100$ однозначно лучше себя показывают параметрические критерии согласия. При рассмотрении объёма выборки $n = 1000$ выяснилось, что параметрические критерии лучше себя показывают лишь в сравнении с Крамером-Мизесом-Смирновым и

Андерсоном-Дарлинггом, но проигрывают в эффективности критериям Колмогорова и Смирнова.

При сравнении непараметрических критериев между собой выяснилось, что в случае рассмотрения сложных гипотез при $n = 1000$ однозначным лидером по эффективности оказался критерий Смирнова. В отношении $n = 50, 100$ все непараметрические критерии ведут себя плюс-минус одинаково.

В ситуации с простыми гипотезами лидером по эффективности при $n = 1000$ всё также остаётся критерий Смирнова, но при $n = 50, 100$ непараметрические критерии вновь ведут себя примерно одинаково.