

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 №6

по дисциплине: Компьютерные технологии моделирования и анализа данных
на тему: Экспериментальное исследование предельных распределений
статистик непараметрических критериев согласия. Часть 2.

Вариант №2

Факультет: ФПМИ

Группа: ПММ-21

Выполнил: Сухих А.С., Черненко Д.А.

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

Дата выполнения: 11.12.22

Отметка о защите:

Новосибирск 2022

Цель работы. Исследование распределений статистик непараметрических критериев согласия при проверке простых и различных сложных гипотез. Во второй части исследуются распределения статистик критериев согласия Купера, Ватсона, Жанга со статистиками Z_K, Z_A, Z_C .

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.

Купер:

$n = 50$	$n = 100$	$n = 1000$
$P = 1 - G(S H_0)$ $= 0.02949820129633391$ ОТВЕРГАЕТСЯ	$P = 1 - G(S H_0)$ $= 0.111650002705322$ НЕ ОТВЕРГАЕТСЯ	$P = 1 - G(S H_0)$ $= 0.464038817060303$ НЕ ОТВЕРГАЕТСЯ

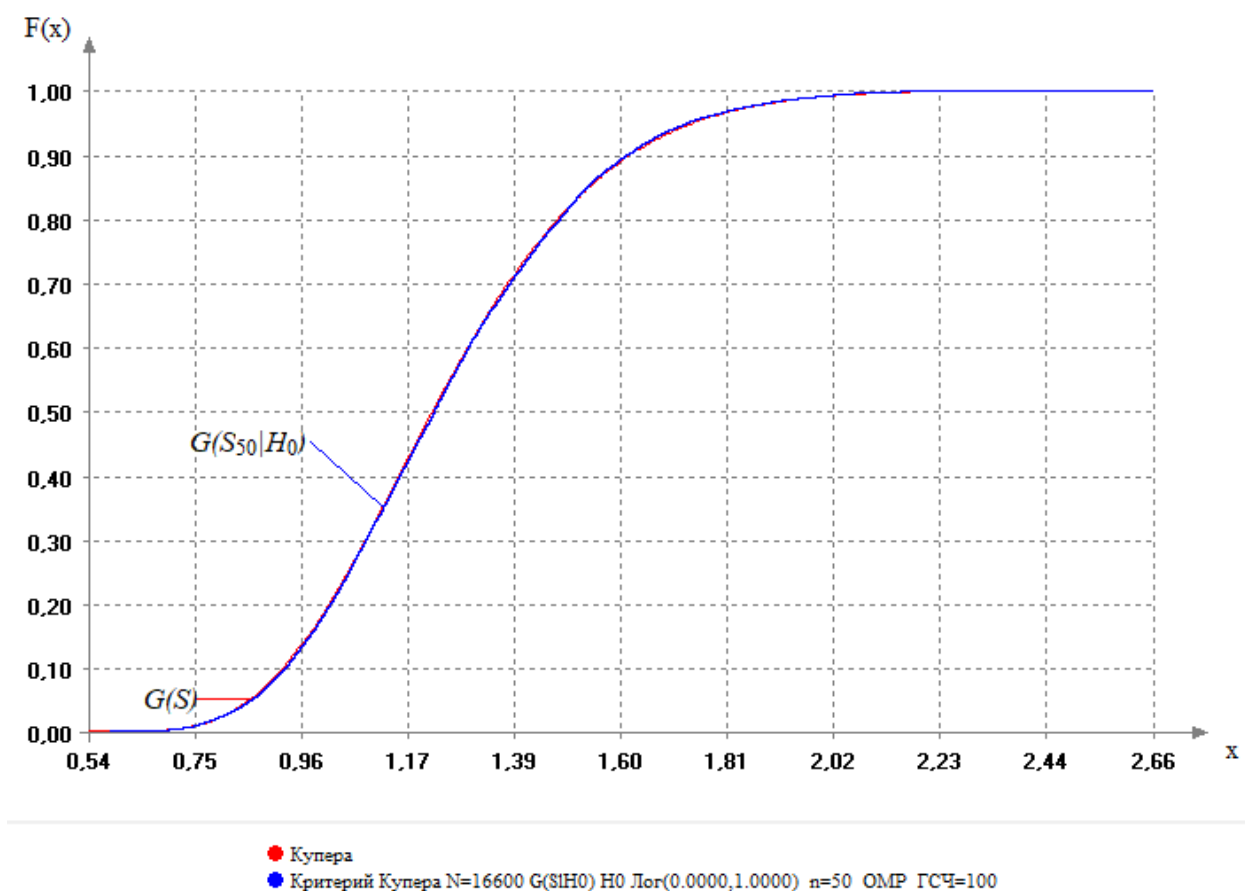


Рисунок 1.1 — сравнение смоделированного распределения Купера с предельным при $n = 50$

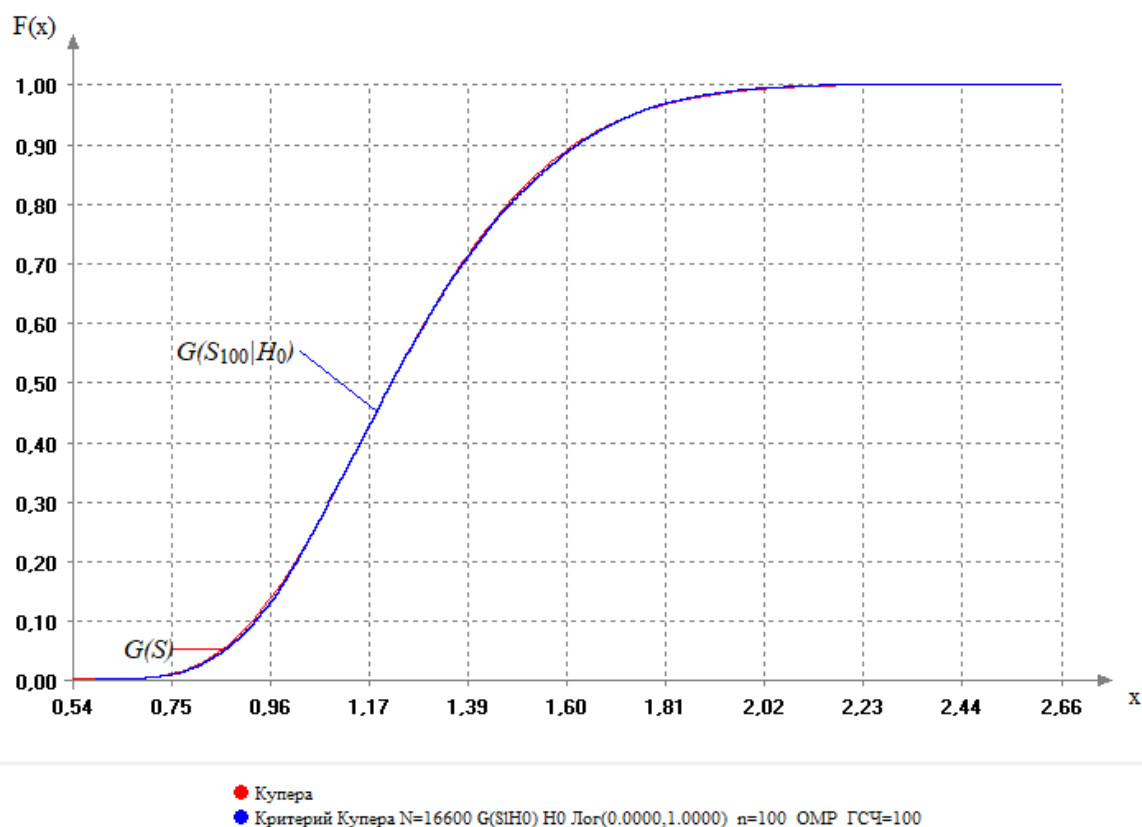


Рисунок 1.2 — сравнение смоделированного распределения Купера с предельным при $n = 100$

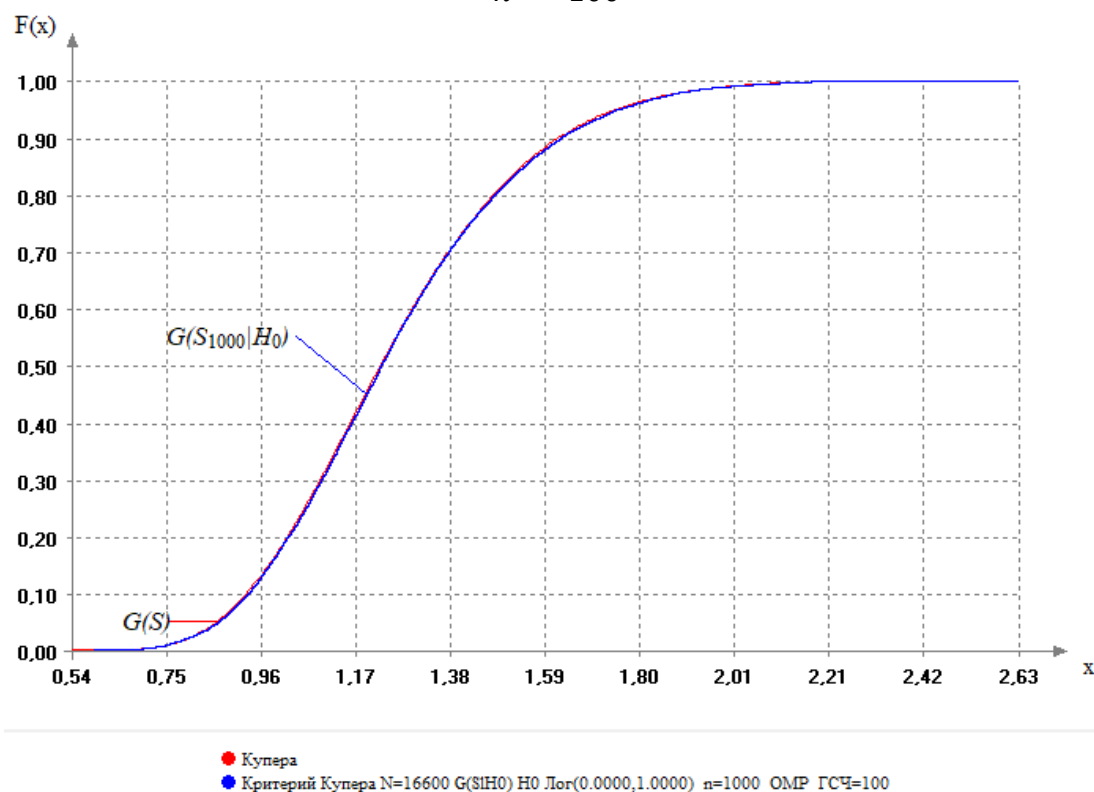
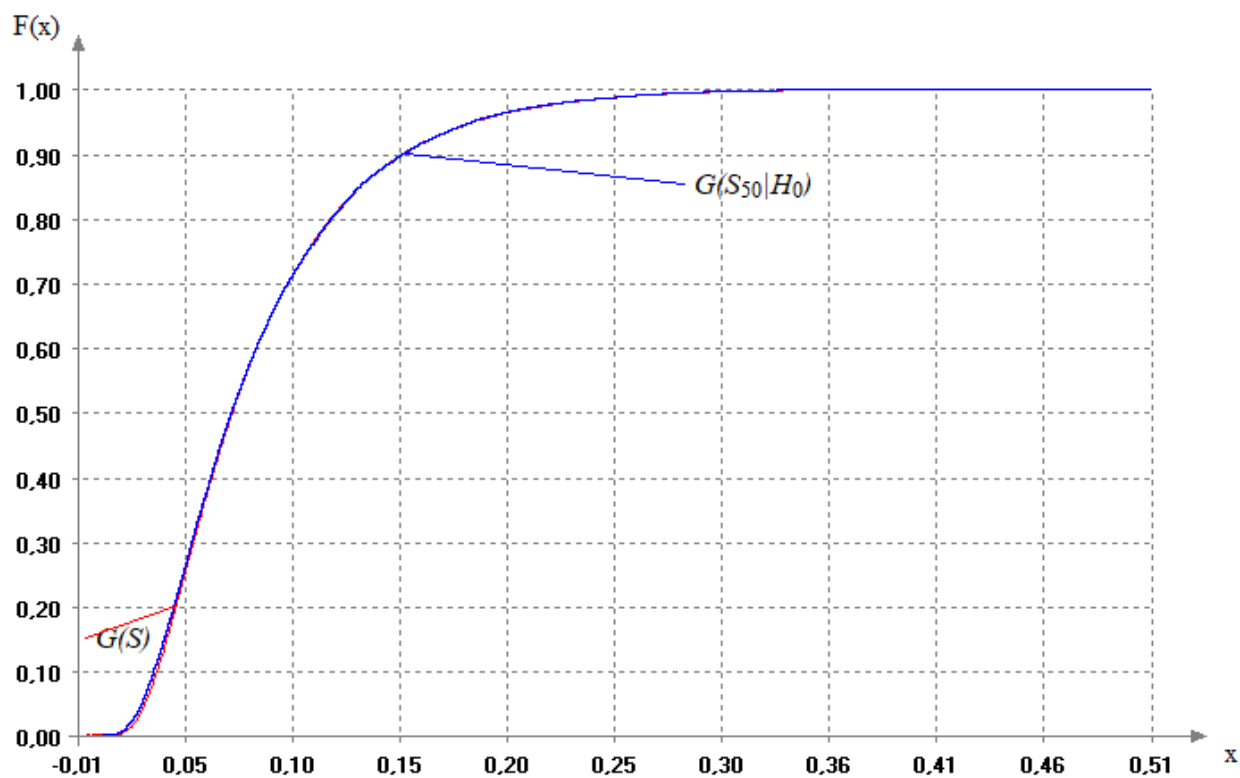


Рисунок 1.3 — сравнение смоделированного распределения Купера с предельным при $n = 1000$

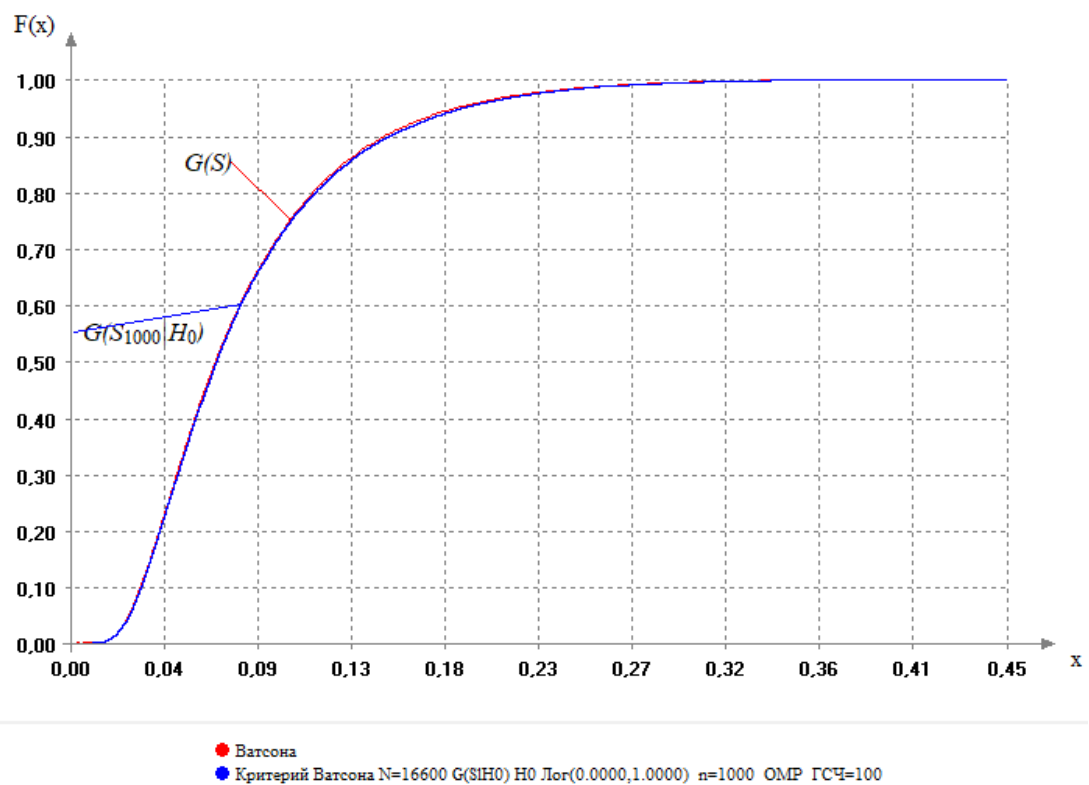
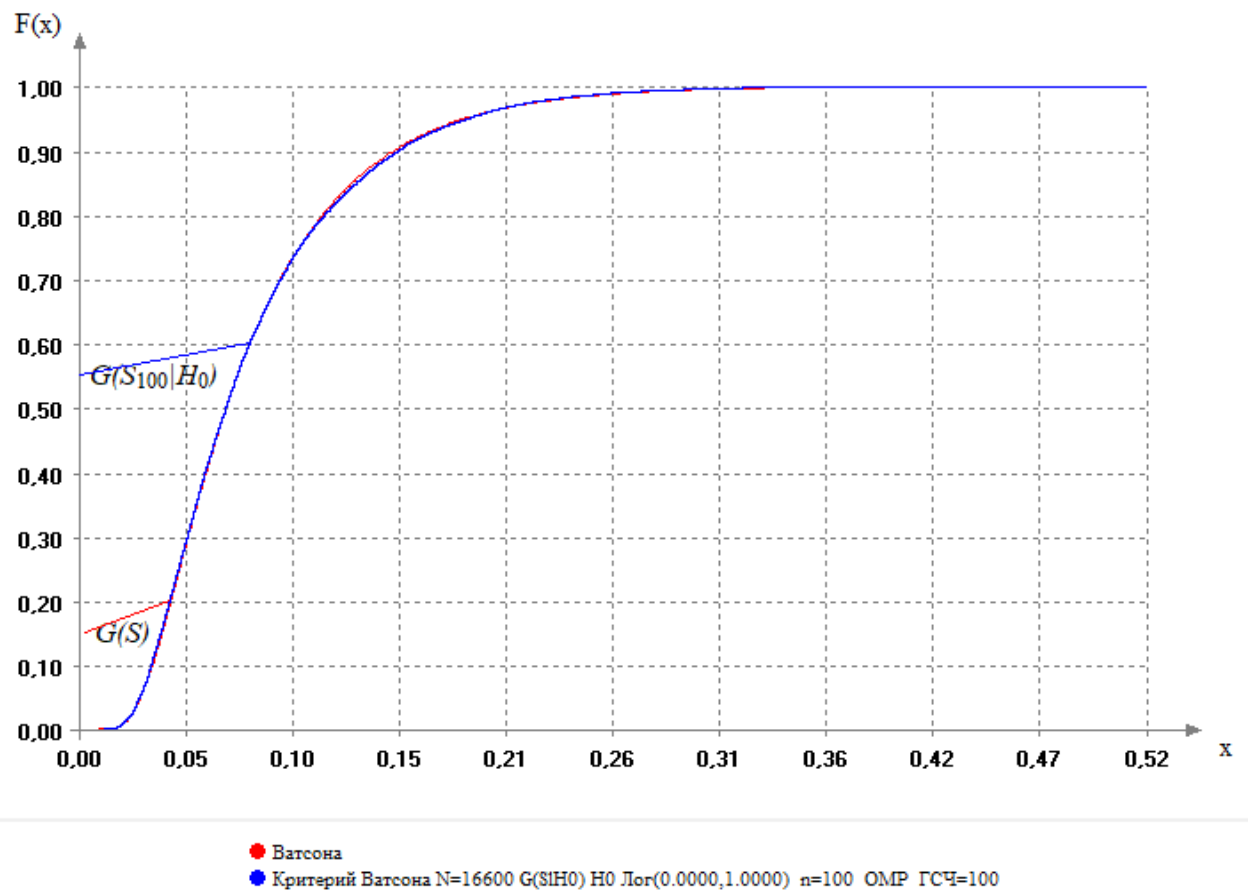
Ватсон:

$n = 50$	$n = 100$	$n = 1000$
$P = 1 - G(S H_0)$ $= 1.812117672549 * 10^{-5}$ ОТВЕРГАЕТСЯ	$P = 1 - G(S H_0)$ $= 0.008932589555653552$ ОТВЕРГАЕТСЯ	$P = 1 - G(S H_0)$ $= 0.1784155569395068$ НЕ ОТВЕРГАЕТСЯ



● Ватсона
● Критерий Ватсона N=16600 G(S|H0) H0 Лог(0.0000,1.0000) n=50 ОМР ГСЧ=100

Рисунок 1.4 — сравнение смоделированного распределения Ватсона с предельным при $n = 50$



Жанг Z_k :

Так как у статистик критериев Жанга нет предельных распределений, а также распределение статистик зависит от объёма выборки, то распределения с объёмами $n = 50, 100, 1000$ я буду сравнивать с $n = 2000$.

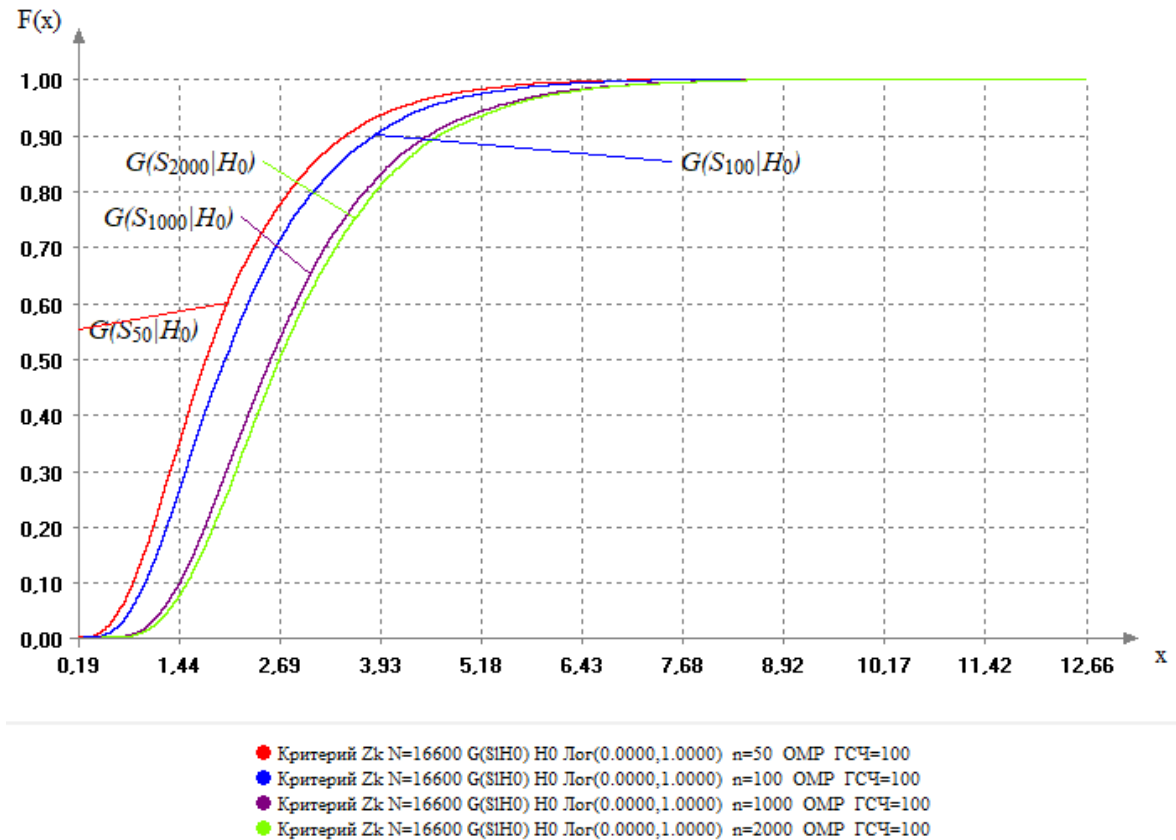


Рисунок 1.7 — сравнение смоделированных распределений Жанга Z_k с распределением с $n = 2000$

К сожалению, проверить на согласие нет возможности, так как isw не позволяет открывать файлы с расширением .dat, которым обладают смоделированные мною выборки.

Жанг Z_a :

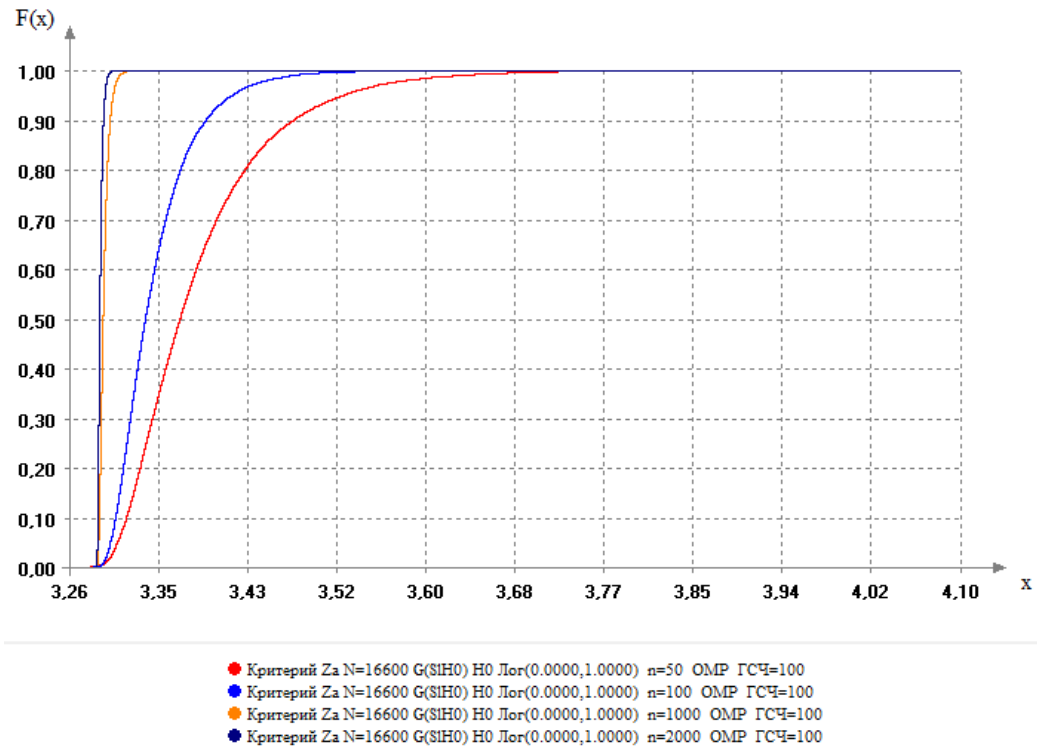


Рисунок 1.8 — сравнение смоделированных распределений Жанга Z_a с распределением с $n = 2000$

Жанг Z_c :

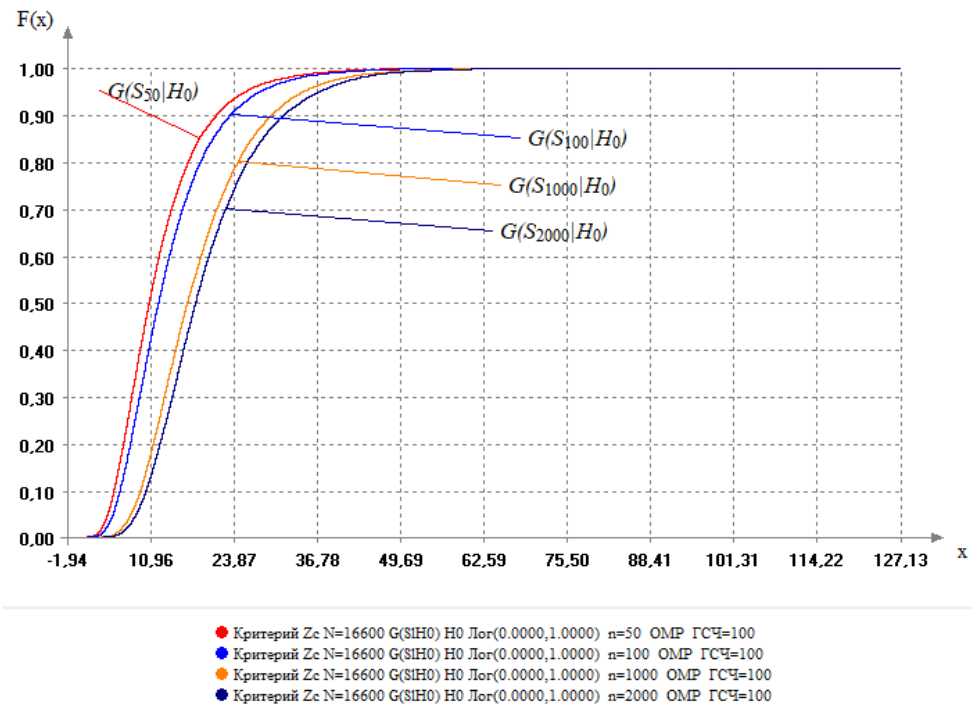


Рисунок 1.9 — сравнение смоделированных распределений Жанга Z_c с распределением с $n = 2000$

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.

Для всех распределений статистик критериев нашей бригадой был выставлен критерий согласия Хи-Квадрат Пирсона, а для критериев согласия Купера и Ватсона мы дополнительно пытались идентифицировать наиболее подходящие распределения по данным критериям согласия, но isw корректно не работала, к сожалению, поэтому этих данных здесь не будет.

Купер:

$n = 50$	$n = 100$	$n = 1000$
Гамма (5.914, 0.083, 0.571)	Бе-III (6.4856, 6.3943, 2.2981, 1.7055, 0.5197)	Гамма (8.5080, 0.0703, 0.4654)
$P = 1 - G(S H_0)$ = 5.628117979117486 * 10^{-6}	$P = 1 - G(S H_0)$ = 0.00278919735649453	$P = 1 - G(S H_0)$ = 0.4370791589141314
ОТВЕРГАЕТСЯ	ОТВЕРГАЕТСЯ	НЕ ОТВЕРГАЕТСЯ

Для $n = 5000$ наиболее подходящим распределением является Гамма (6.1160, 0.0858, 0.5392).

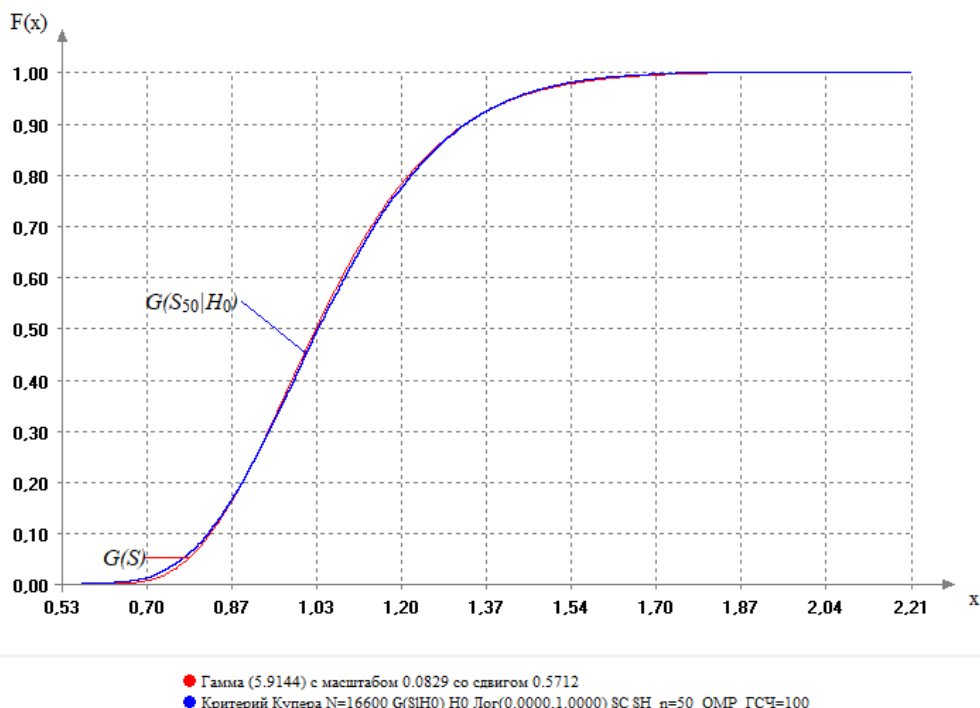


Рисунок 2.1 — Сопоставление смоделированного распределения Купера $n = 50$ с Гамма распределением

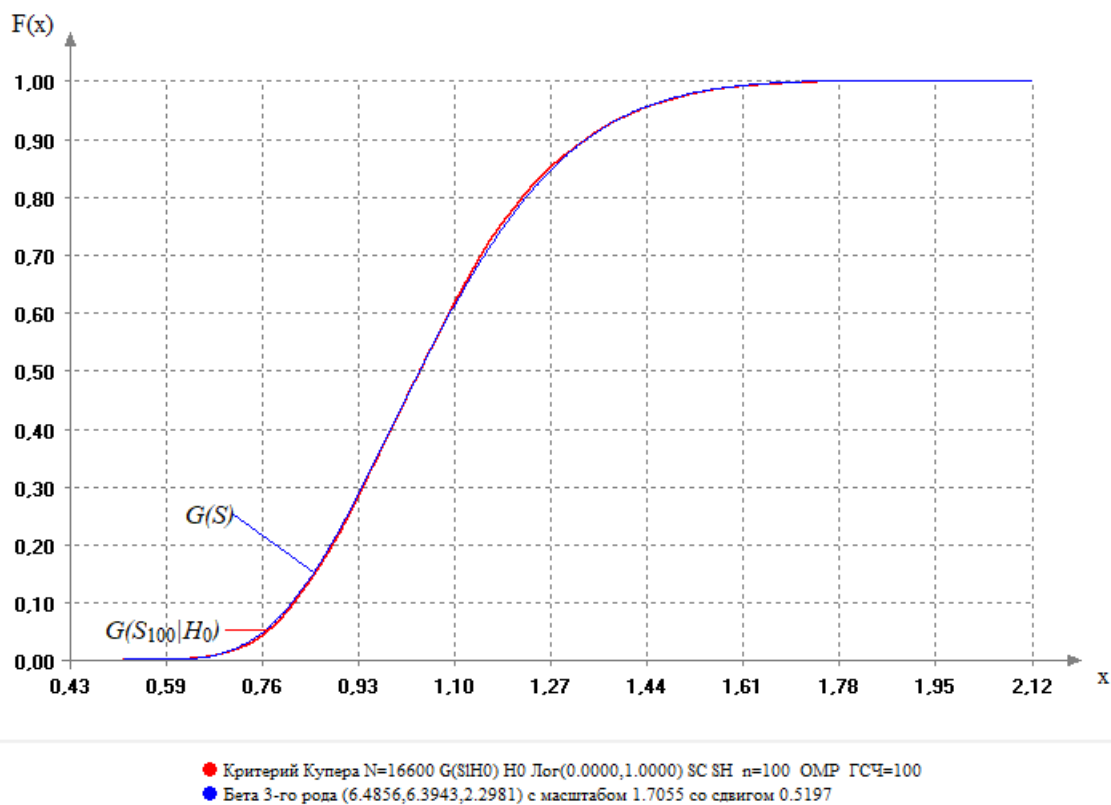


Рисунок 2.2 — Сопоставление смоделированного распределения Купера $n = 100$ с Бета-III распределением

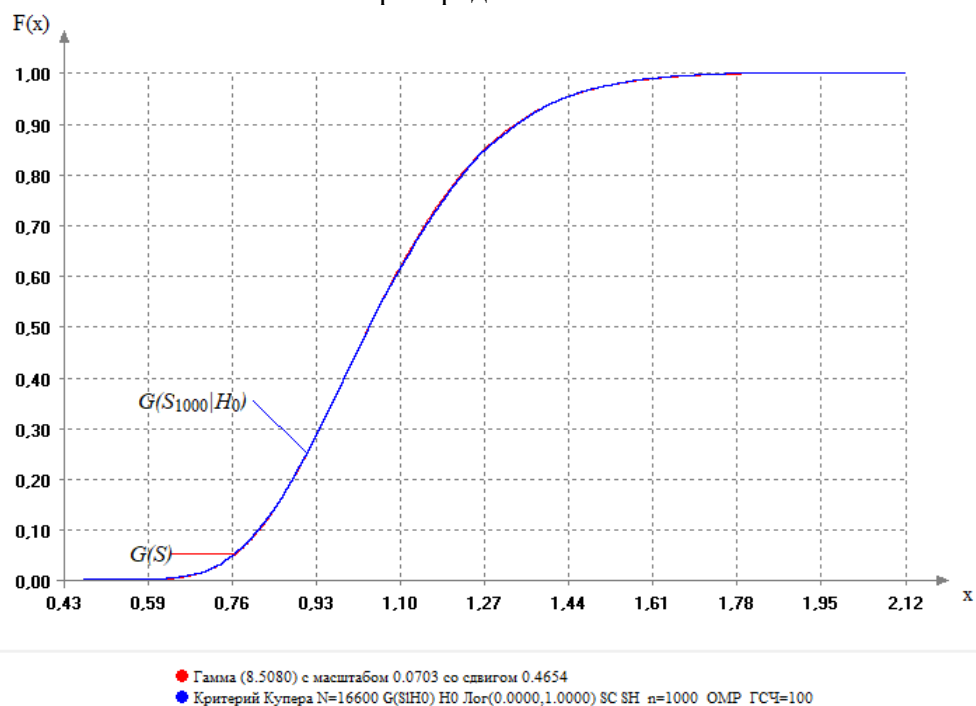
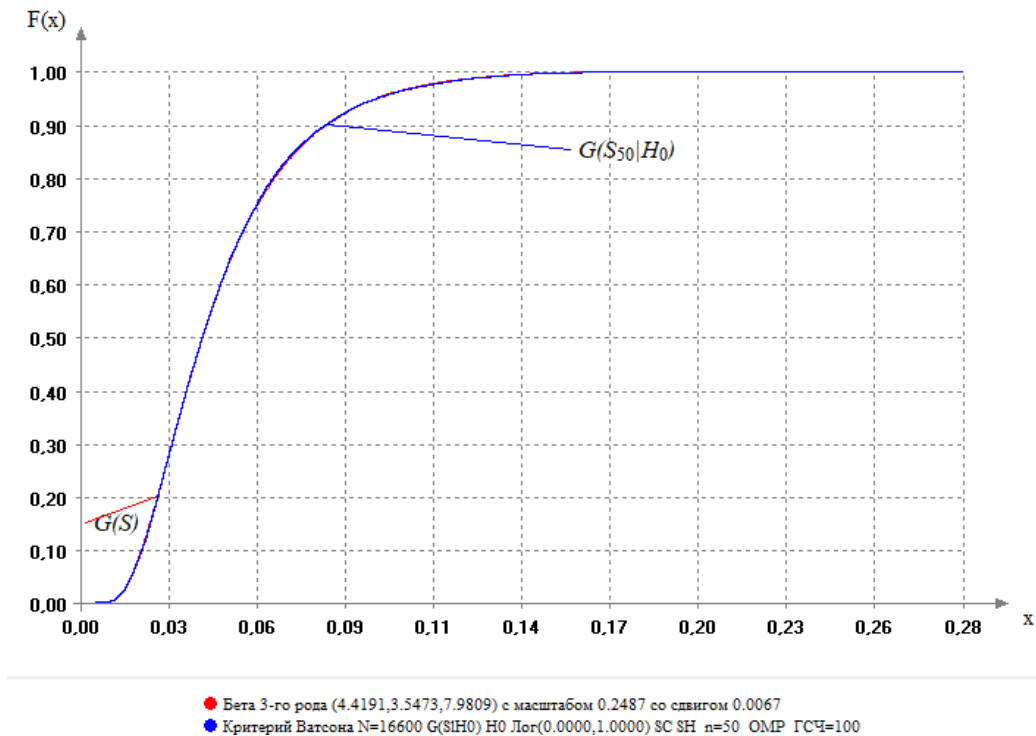


Рисунок 2.3 — Сопоставление смоделированного распределения Купера $n = 1000$ с Гамма распределением

Ватсон:

$n = 50$	$n = 100$	$n = 1000$
Бе-III (4.4115, 3.5511, 7.9560, 0.2487, 0.0067)	Бе-III (3.8900, 4.0724, 6.7965, 0.2722, 0.0084)	Бе-III (5.618, 3.485, 9.758, 0.2445, 0.0067)
$P = 1 - G(S H_0)$ $= 0.2038690115925457$ НЕ ОТВЕРГАЕТСЯ	$P = 1 - G(S H_0)$ $= 0.09575831514776979$ ОТВЕРГАЕТСЯ	$P = 1 - G(S H_0)$ $= 0.1301202927952418$ НЕ ОТВЕРГАЕТСЯ

Для $n = 5000$ наиболее подходящим распределением является Бе-III (5.4507, 4.1856, 10.9990, 0.3321, 0.0066).



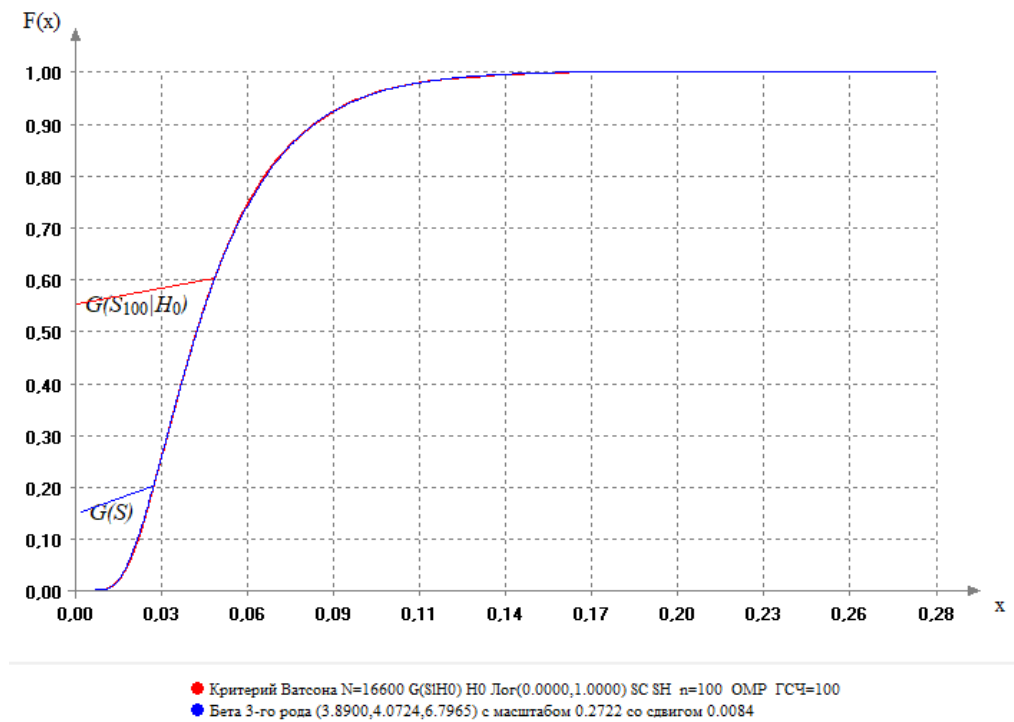


Рисунок 2.5 — Сопоставление смоделированного распределения Ватсона $n = 100$ с Бета-III распределением

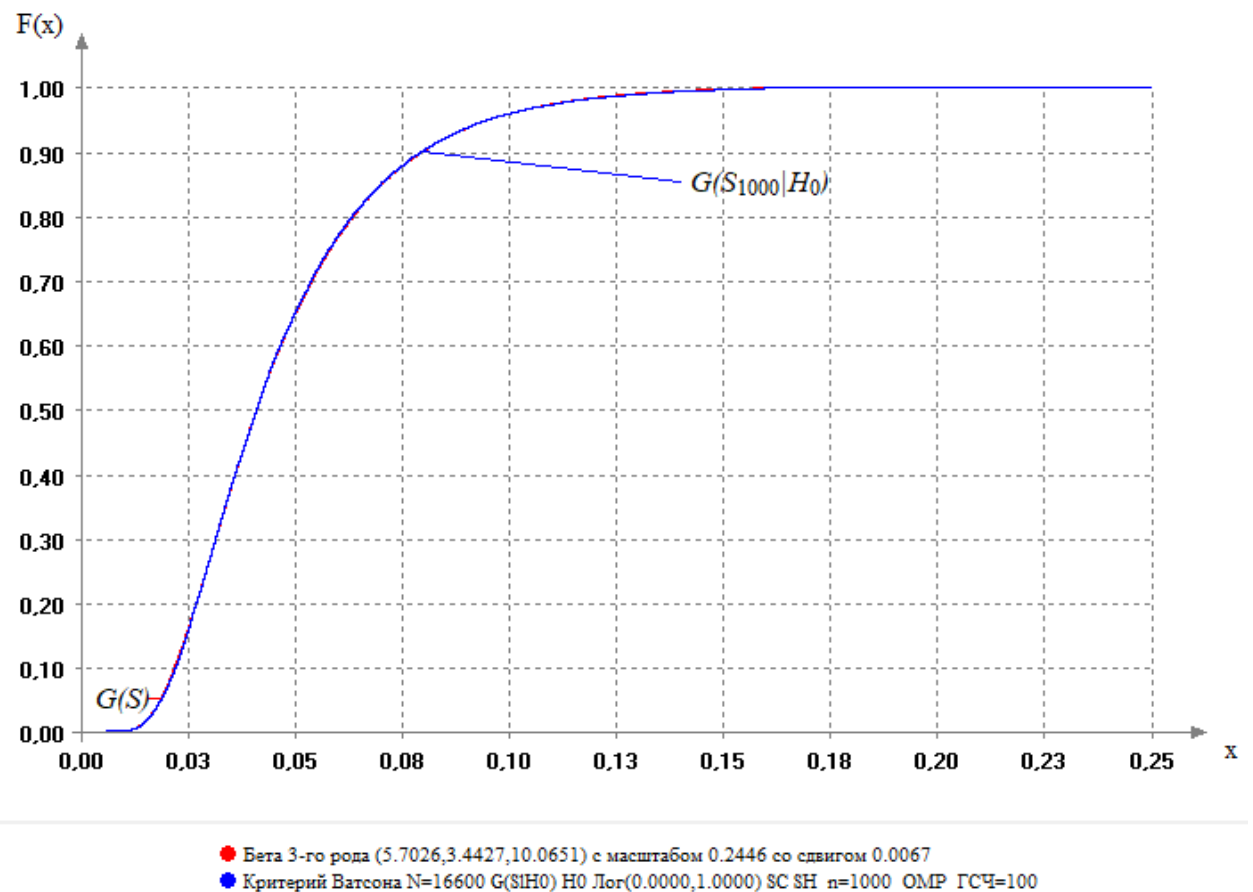
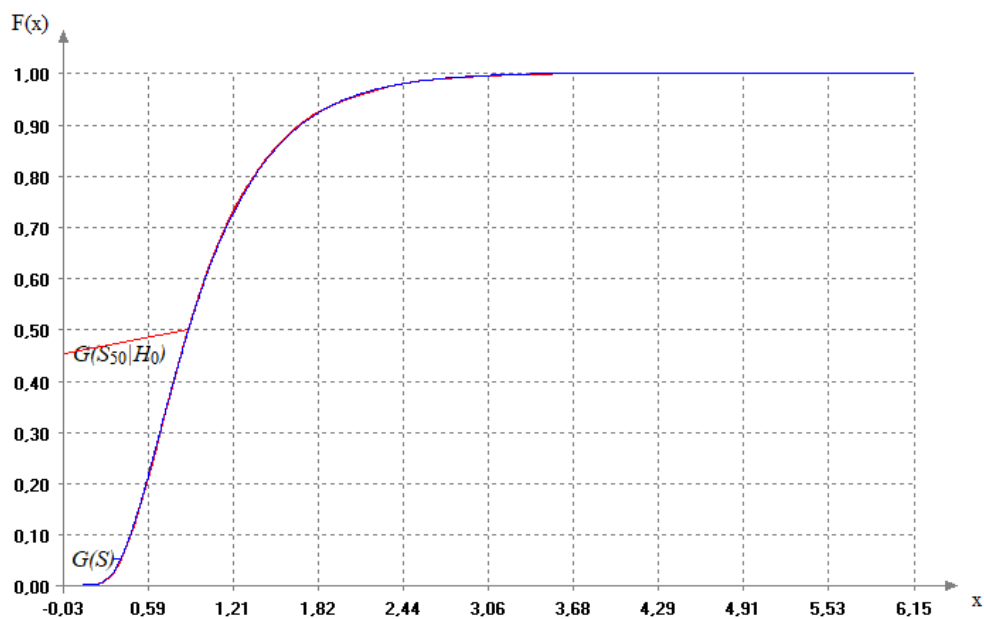


Рисунок 2.6 — Сопоставление смоделированного распределения Ватсона $n = 1000$ с Бета-III распределением

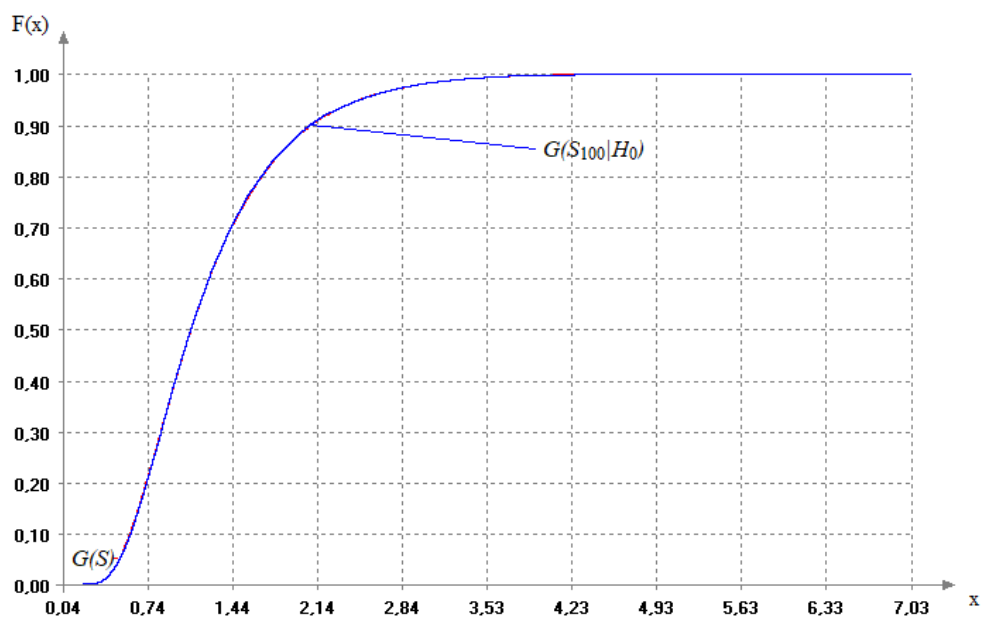
Жанг Z_k :

$n = 50$	$n = 100$	$n = 1000$
Бе-III (4.9230, 4.1842, 8.0262, 5.9528, 0.1179)	Бе-III (4.3781, 3.9749, 7.2539, 6.7300, 0.2094)	Sb-Дж (3.0679, 1.4903, 12.3442, 0.3630)
$P = 1 - G(S H_0)$ = 0.1398338009829733 НЕ ОТВЕРГАЕТСЯ	$P = 1 - G(S H_0) =$ 0.09674356367879455 ОТВЕРГАЕТСЯ	$P = 1 - G(S H_0)$ = 0.006883821769333062 ОТВЕРГАЕТСЯ



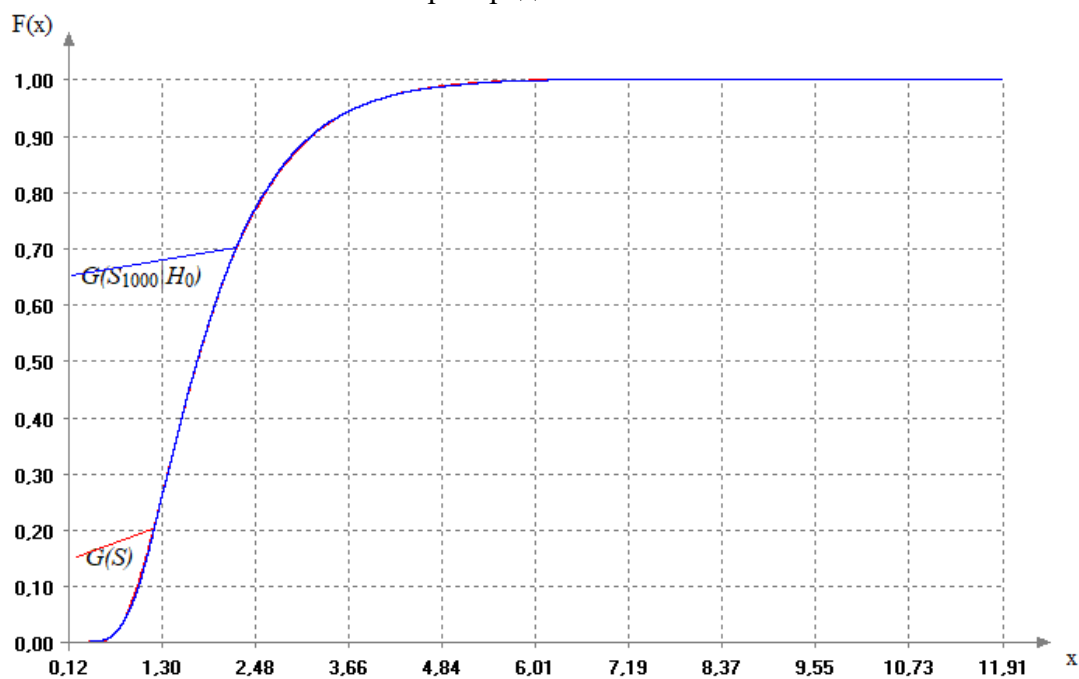
● Критерий Z_k $N=16600$ $G(S|H_0)$ H_0 Лор(0.0000, 1.0000) SC SH $n=50$ ОМР ГСЧ=100
● Бета 3-го рода (4.9230, 4.1842, 8.0262) с масштабом 5.9528 со сдвигом 0.1179

Рисунок 2.7 — Сопоставление смоделированного распределения Z_k $n = 50$ с Бета-III распределением



● Бета 3-го рода (4.3781,3.9749,7.2539) с масштабом 6.7300 со сдвигом 0.2094
 ● Критерий Z_k $N=16600$ $G(S|H_0)$ H_0 $\text{Log}(0.0000,1.0000)$ SC SH $n=100$ OMP $ГСЧ=100$

Рисунок 2.8 — Сопоставление смоделированного распределения Z_k $n = 100$ с Бета-III распределением

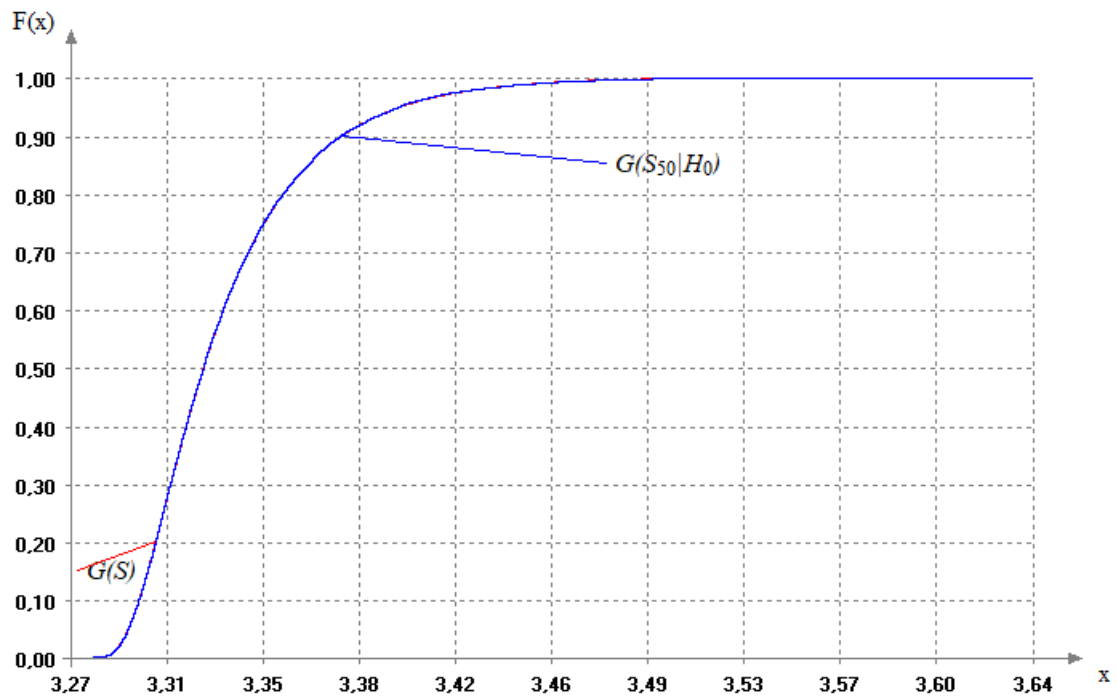


● Су-Джонсона (3.0679,1.4903) с масштабом 12.3442 со сдвигом 0.3630
 ● Критерий Z_k $N=16600$ $G(S|H_0)$ H_0 $\text{Log}(0.0000,1.0000)$ SC SH $n=1000$ OMP $ГСЧ=100$

Рисунок 2.9 — Сопоставление смоделированного распределения Z_k $n = 1000$ с распределением Су-Джонсона

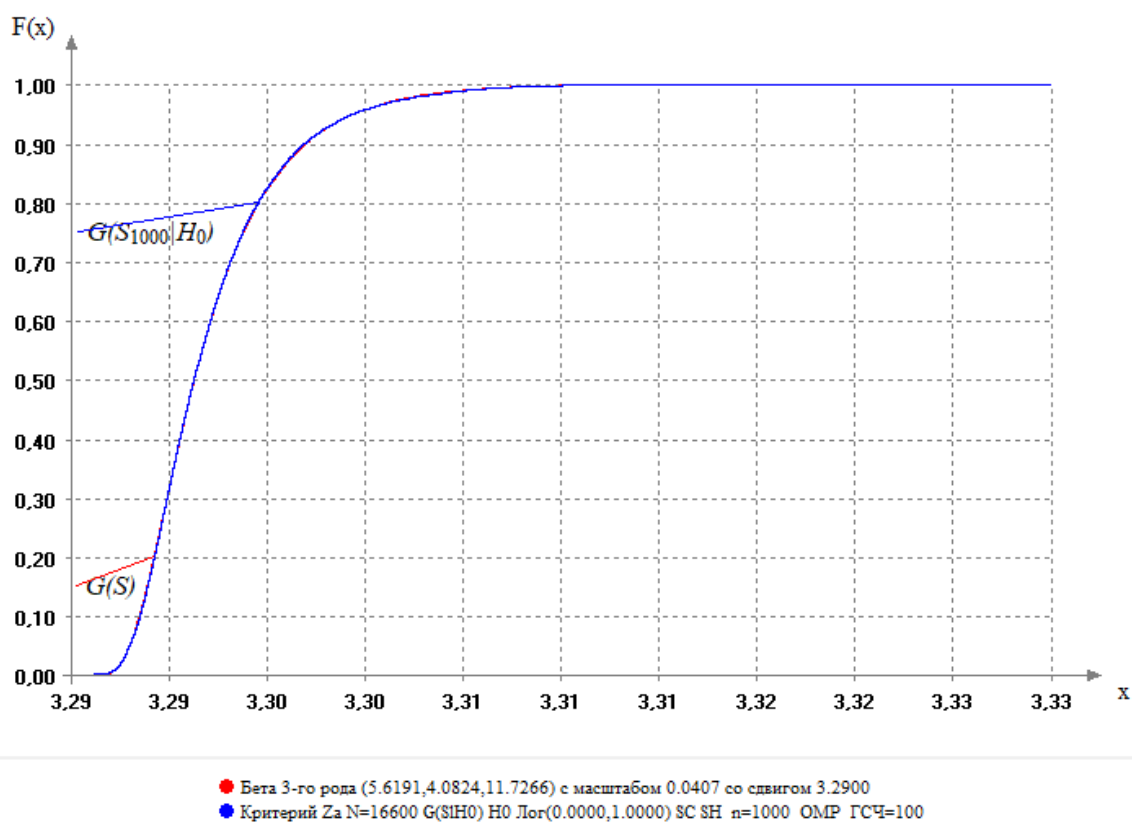
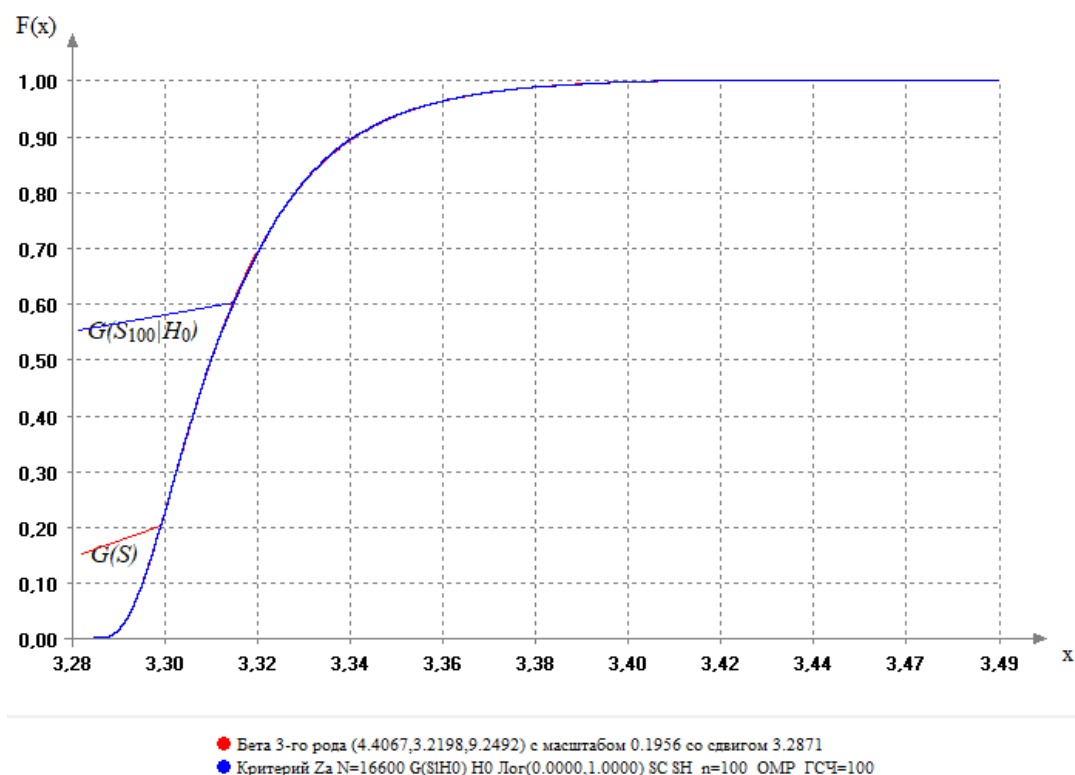
Жанг Z_a :

$n = 50$	$n = 100$	$n = 1000$
Бе-III (4.3133, 3.3829, 9.6243, 0.3538, 3.2812)	Бе-III (4.4067, 3.2198, 9.2492, 0.1956, 3.2871)	Бе-III (5.6191, 4.0824, 11.7266, 0.0407, 3.2900)
$P = 1 - G(S H_0)$ = 0.5001289996283291 НЕ ОТВЕРГАЕТСЯ	$P = 1 - G(S H_0)$ = 0.04462262267054866 ОТВЕРГАЕТСЯ	$P = 1 - G(S H_0)$ = 0.1256480132025264 НЕ ОТВЕРГАЕТСЯ



● Бета 3-го рода (4.3133, 3.3829, 9.6243) с масштабом 0.3538 со сдвигом 3.2812
● Критерий Z_a $N=16600$ $G(S|H_0)$ H_0 $\text{Log}(0.0000, 1.0000)$ SC SH $n=50$ OMP $ГСЧ=100$

Рисунок 2.10 — Сопоставление смоделированного распределения Z_a $n = 50$ с распределением Бе-III



Жанг Z_c :

$n = 50$	$n = 100$	$n = 1000$
Su-Дж (-3.7396, 1.8334, 1.4687, 0.5677)	Бе-П (5.2121, 9.4328, 12.3935, 0.8145)	Su-Дж (-3.9127, 1.9731, 2.8087, 1.8116)
$P = 1 - G(S H_0)$ = 5.471785924367941 $\cdot 10^{-5}$ ОТВЕРГАЕТСЯ	$P = 1 - G(S H_0)$ = 0.04179641648362855 ОТВЕРГАЕТСЯ	$P = 1 - G(S H_0)$ = 0.1634849979102518 НЕ ОТВЕРГАЕТСЯ

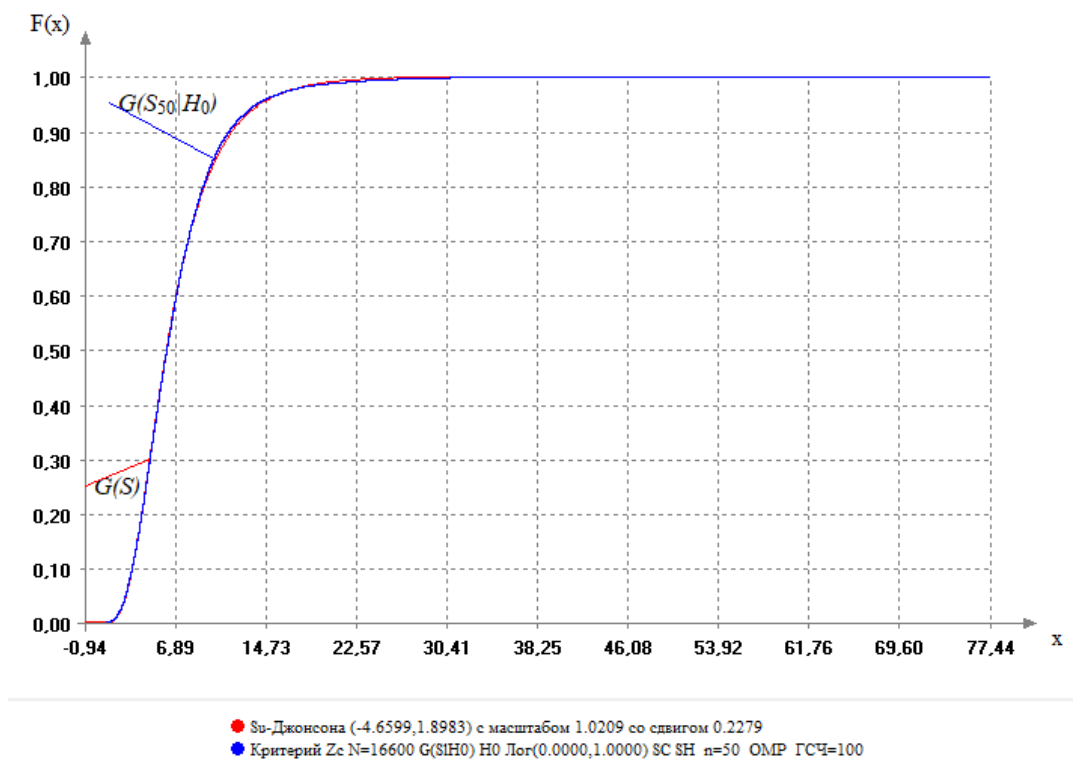


Рисунок 2.13 — Сопоставление смоделированного распределения Z_c $n = 50$ с распределением Су-Джонсона

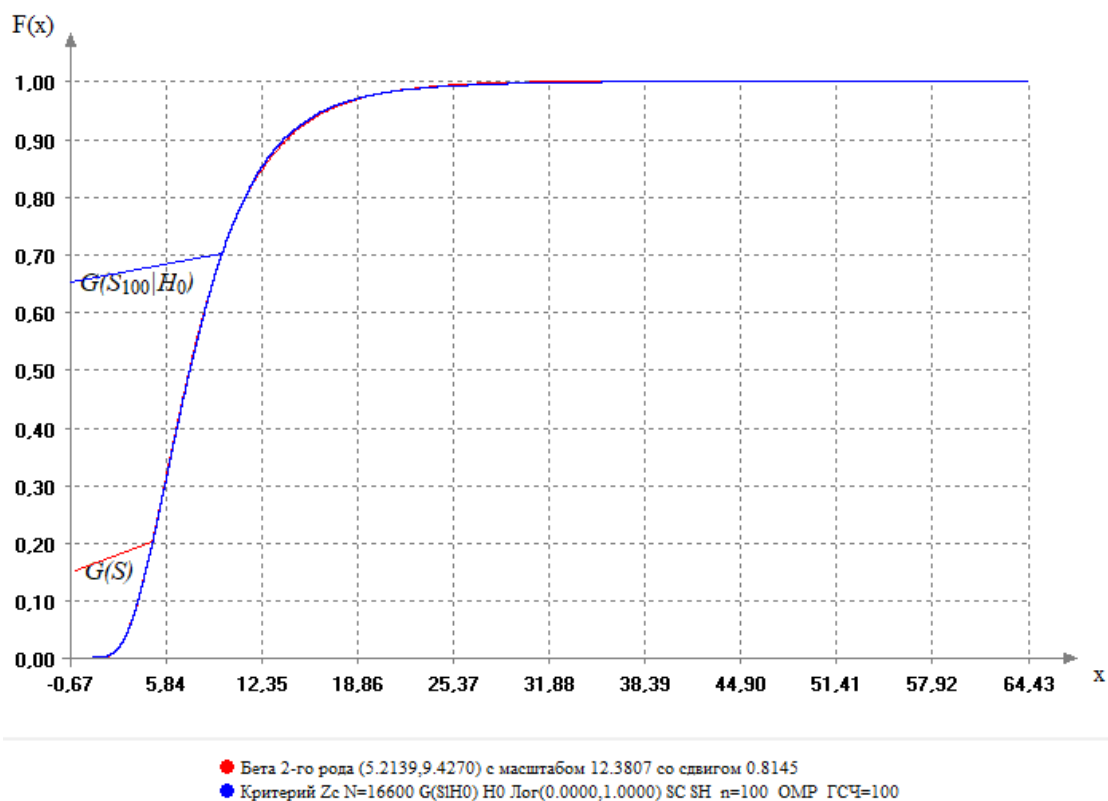


Рисунок 2.14 — Сопоставление смоделированного распределения Z за $n = 100$ с распределением Бе-П

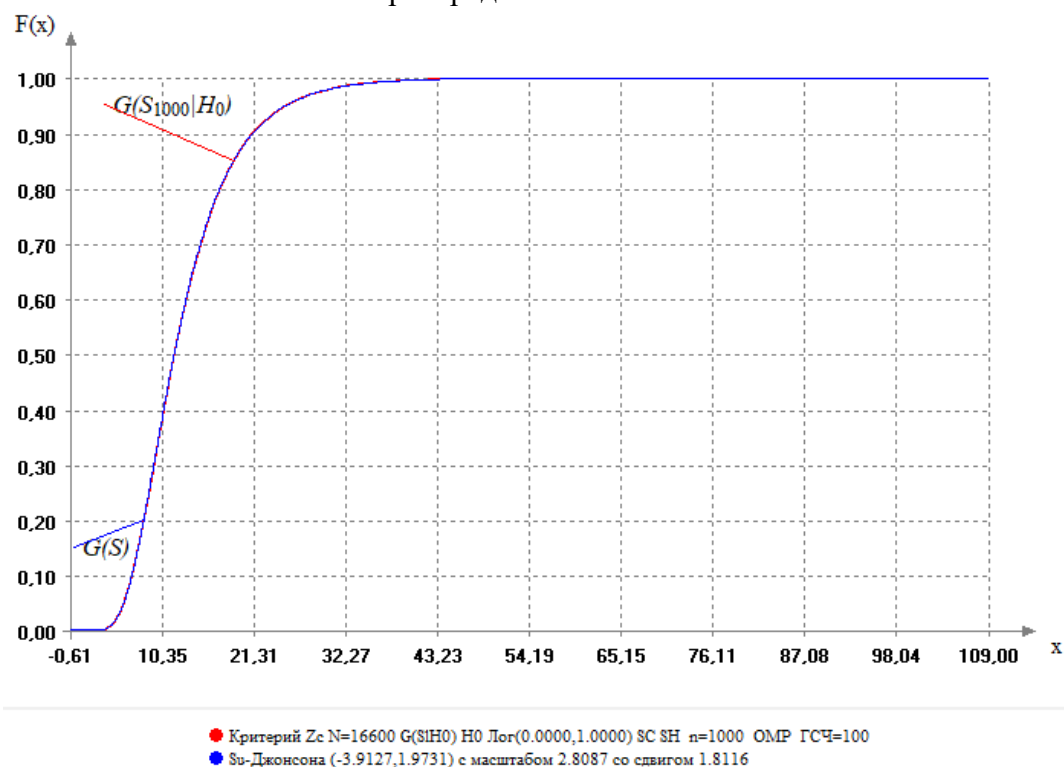


Рисунок 2.15 — Сопоставление смоделированного распределения Z за $n = 1000$ с распределением Су-Джонсона

Дополнительно были исследованы распределения статистик Купера и Ватсона $n = 5000$. Выяснилось, что Критерию Купера лучше всего подходит статистика Гамма ($P = 3.28948951032471e-07$), гипотеза о согласии в данном случае **отвергается**. Критерию Ватсона всё также подходит статистика Бета-III ($P = 0.4215775845806789$, **не отвергается**). Данные результаты подтверждают теоретические сведения о сильной зависимости критерия Купера от объёма выборки (т.к. при её увеличении наиболее похожим распределением является Гамма, а не Бета-III), критерий Ватсона же напротив, практически не зависит от объёма выборки, и даже при $n = 5000$ наиболее подходящим распределением для него остаётся Бета-III.

Для критериев Жанга всё выглядит достаточно сумбурно. Как мы знаем из теории, статистика критериев Z_k, Z_a, Z_c зависит от n . Для статистики Z_c , по всей видимости, при повышении объёма выборки, повышается вероятность согласия с проверяющим распределением (Su-Джонсон). Критериям Z_a и Z_k достаточно хорошо подходит распределение Бета-III, а также при повышении n Z_k начинает хуже согласовываться с распределениями Бета-III и Sb-Джонсона, а у Z_a провал в возможности согласования возникает лишь при $n = 100$.

Как мы можем увидеть по рисункам, у нас графики всех распределений накладываются друг на друга, но, к сожалению, это ещё не означает, что все проверяемые распределения согласуются с подобранными для проверки на согласие.

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 мощности всех непараметрических критериев.

Для Купера (V_n):

α	Простая гипотеза			Сложная гипотеза		
	$1 - \beta$					
	$n = 50$	$n = 100$	$n = 1000$	$n = 50$	$n = 100$	$n = 1000$
0.15	0.215723	0.263855	0.882289	0.193373	0.240422	0.86494
0.1	0.15259	0.198795	0.836145	0.135181	0.17247	0.800422
0.05	0.0881928	0.118675	0.749759	0.0689157	0.0987952	0.685301
0.025	0.0460241	0.0727711	0.650723	0.0363253	0.0560843	0.555241
0.01	0.021506	0.0390361	0.511205	0.0150602	0.026506	0.405542

Для Ватсона (U_n^2):

α	Простая гипотеза			Сложная гипотеза		
	$1 - \beta$					
	$n = 50$	$n = 100$	$n = 1000$	$n = 50$	$n = 100$	$n = 1000$
0.15	0.218313	0.272289	0.90259	0.19741	0.254337	0.918976
0.1	0.154277	0.202651	0.860241	0.133614	0.184819	0.872651
0.05	0.0898193	0.124578	0.780964	0.0693373	0.103795	0.77759
0.025	0.0496386	0.078253	0.687048	0.035	0.0613855	0.660904
0.01	0.0233735	0.0413253	0.562108	0.016506	0.0254217	0.516928

Для Жанга Z_k :

α	Простая гипотеза			Сложная гипотеза		
	$1 - \beta$					
	$n = 50$	$n = 100$	$n = 1000$	$n = 50$	$n = 100$	$n = 1000$
0.15	0.170181	0.189639	0.755422	0.126084	0.161386	0.93006
0.1	0.121506	0.134036	0.64494	0.0766867	0.103373	0.875542
0.05	0.0677711	0.0754217	0.460241	0.0326506	0.0448795	0.752048
0.025	0.0369277	0.0398795	0.324518	0.015	0.0204819	0.613434
0.01	0.0160241	0.0180723	0.178012	0.00427711	0.00740964	0.415181

Для Жанга Z_a :

α	Простая гипотеза			Сложная гипотеза		
	$1 - \beta$					
	$n = 50$	$n = 100$	$n = 1000$	$n = 50$	$n = 100$	$n = 1000$
0.15	0.114398	0.147048	0.958675	0.156807	0.269036	0.995181
0.1	0.0706024	0.0906627	0.919337	0.102048	0.195181	0.990783
0.05	0.0327108	0.0389759	0.826205	0.0492169	0.102048	0.974578
0.025	0.0162048	0.016747	0.702048	0.021506	0.0540964	0.936988
0.01	0.00578313	0.00620482	0.528313	0.00722892	0.0226506	0.866687

Для Жанга Z_c :

α	Простая гипотеза			Сложная гипотеза		
	$1 - \beta$					
	$n = 50$	$n = 100$	$n = 1000$	$n = 50$	$n = 100$	$n = 1000$
0.15	0.116265	0.144398	0.956928	0.176928	0.286024	0.995482
0.1	0.0694578	0.0859036	0.906446	0.103193	0.18494	0.99012
0.05	0.0293373	0.0355422	0.788434	0.0357229	0.070241	0.967892
0.025	0.0122289	0.0137952	0.610602	0.00795181	0.0211446	0.91
0.01	0.00427711	0.0039759	0.361627	0.000240964	0.00186747	0.78747

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:

В ходе выполнения лабораторной работы выяснилось, что иерархия мощностей критериев при простых и сложных гипотезах выглядит следующим образом:

Сложная гипотеза:

$$Z_k > Z_c > Z_a \geq Smir > Kol > Nickul > Pirson > AD > V_n \geq KMS \geq U_n^2$$

Простая гипотеза:

$$Z_c > Z_a > Z_k > Smir > KMS > K > AD > Nickul > Pirson > V_n > U_n^2$$