

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 №7**

по дисциплине: Компьютерные технологии моделирования и анализа данных  
на тему: Исследование критериев проверки отклонения от нормального  
закона. 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: 22.12.22

Отметка о защите:

Новосибирск 2022

## Оглавление

1. Исследовать зависимость распределений статистик критерия симметричности и критерия эксцесса от объема выборок в случае принадлежности наблюдений нормальному закону. При некотором объеме выборок смоделировать эмпирические распределения статистик критерия симметричности и критерия эксцесса при законе двустороннего экспоненциального распределения. Сравнить распределения статистик со случаем нормальности наблюдаемого закона.....	3
2. Исследовать распределения статистик критерия Шапиро-Уилка при различных объемах выборок. Оценить близость получаемых эмпирических распределений статистик к «теоретическим» по процентным точкам таблиц, соответствующим данному критерию. При некотором объеме выборок ( $n=10$ ) смоделировать распределение статистики критерия при обобщённом нормальном законе (двустороннее экспоненциальное) при параметре формы, равном $4 \div 7$ . Сравнить с ситуацией, соответствующей справедливой проверяемой гипотезе о нормальном законе. Оценить мощность критерия. ....	15
3. Аналогично предыдущему пункту исследовать распределения статистики критерия Эппса-Пулли. ....	17
4. Исследовать распределения статистики $z_1$ . Проверить близость эмпирических распределений статистики стандартному нормальному закону. ....	20
5. Исследовать распределения статистики $z_2$ . Проверить близость эмпирических распределений статистики стандартному нормальному закону. ....	25
6. Оценить мощность критериев со статистиками (7.3), (7.4), (7.5), (7.6) относительно заданной альтернативы. ....	30

**Цель работы:** Исследование распределений статистик критериев, используемых при проверке отклонения эмпирических распределений наблюдаемых величин от нормального закона (в том числе критериев проверки гипотез о симметричности и о значении эксцесса при различных наблюдаемых законах). Исследование распределений статистик критериев Шапиро–Уилка, Эппса–Палли, Д’Агостино. Исследование и сравнение мощности критериев относительно близких конкурирующих гипотез.

№ п/п	$F_1(x, \theta_1)$
2	Двустороннее экспоненциальное с параметром формы 5

**1. Исследовать зависимость распределений статистик критерия симметричности и критерия эксцесса от объема выборок в случае принадлежности наблюдений нормальному закону. При некотором объеме выборок смоделировать эмпирические распределения статистик критерия симметричности и критерия эксцесса при законе двустороннего экспоненциального распределения. Сравнить распределения статистик со случаем нормальности наблюдаемого закона.**

По объемам выборок  $n = 50, 100, 1000$  были смоделированы статистики критериев симметричности и эксцесса в случае принадлежности нормальному закону.

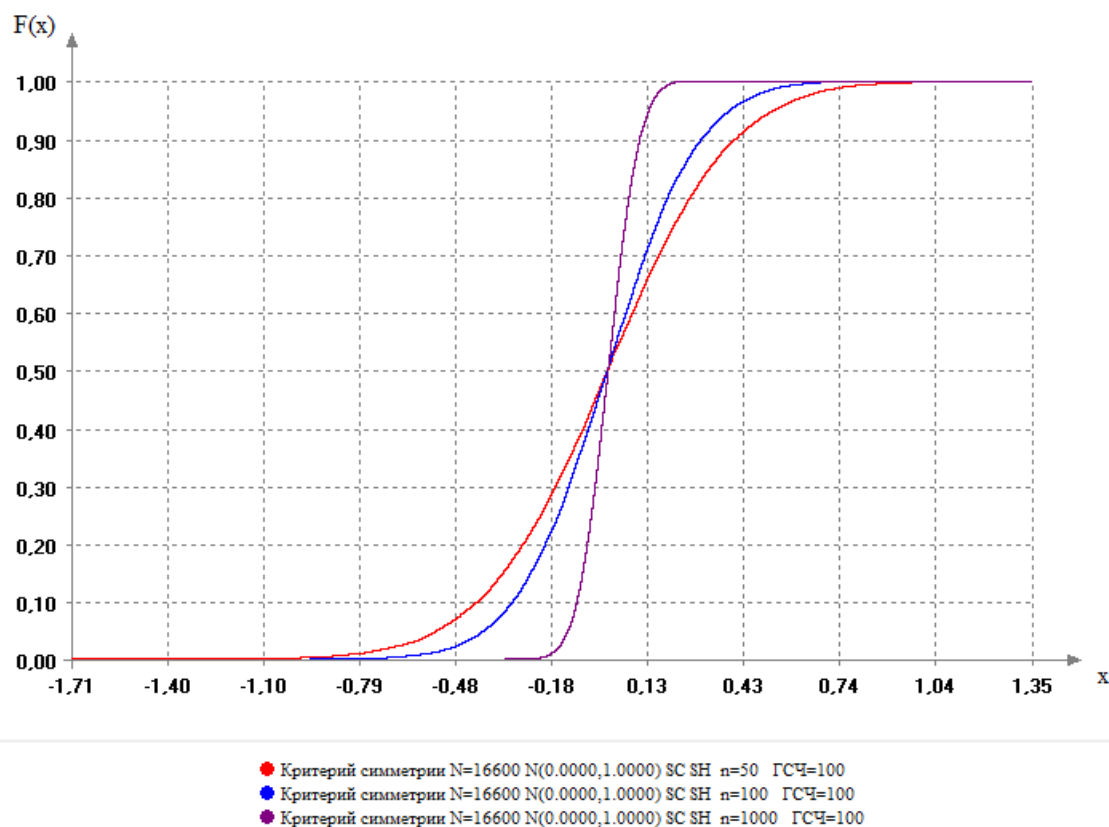


Рисунок 1.1 — Изменение статистики критерия симметричности для нормального закона с изменением объёма выборки

Как можно увидеть, при увеличении объёма выборки статистика критерия симметричности задирается вверх.

Посмотрим на согласие данной статистики с нормальным законом распределения ( $\alpha = 0.1$ ):

$n = 50$	$n = 100$	$n = 1000$
$P = 1.80635482239e - 06$ <b>ОТВЕРГАЕТСЯ</b>	$P = 0.0001735592771341$ <b>ОТВЕРГАЕТСЯ</b>	$P = 0.4658354686747758$ <b>НЕ ОТВЕРГАЕТСЯ</b>

Как можно увидеть, критерий симметричности ведёт себя не столь предсказуемо, как можно было бы подумать, что совпадает с теоретическими данными, так как известно, что критерий симметричности не всегда согласуется с нормальным законом распределения (т.к. он может обоснованно отклоняться и неотклоняться).

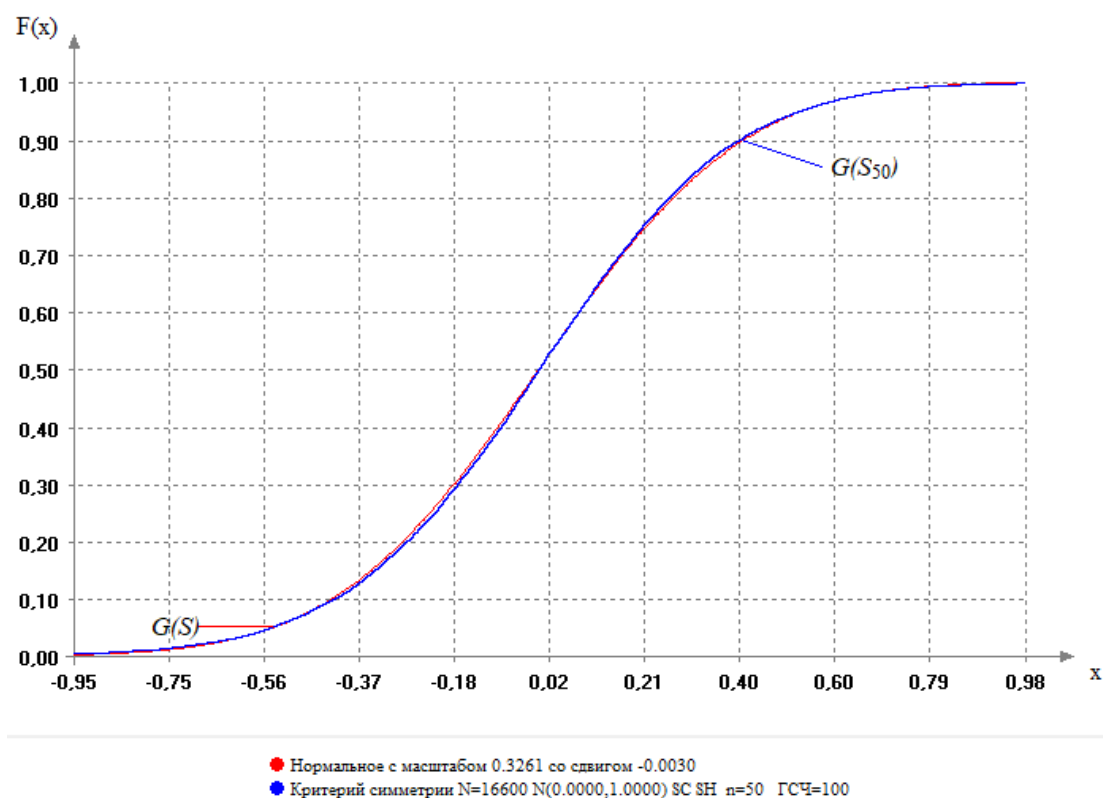


Рисунок 1.2 — Проверка на согласие критерия симметричности при  $n = 50$  с нормальным распределением

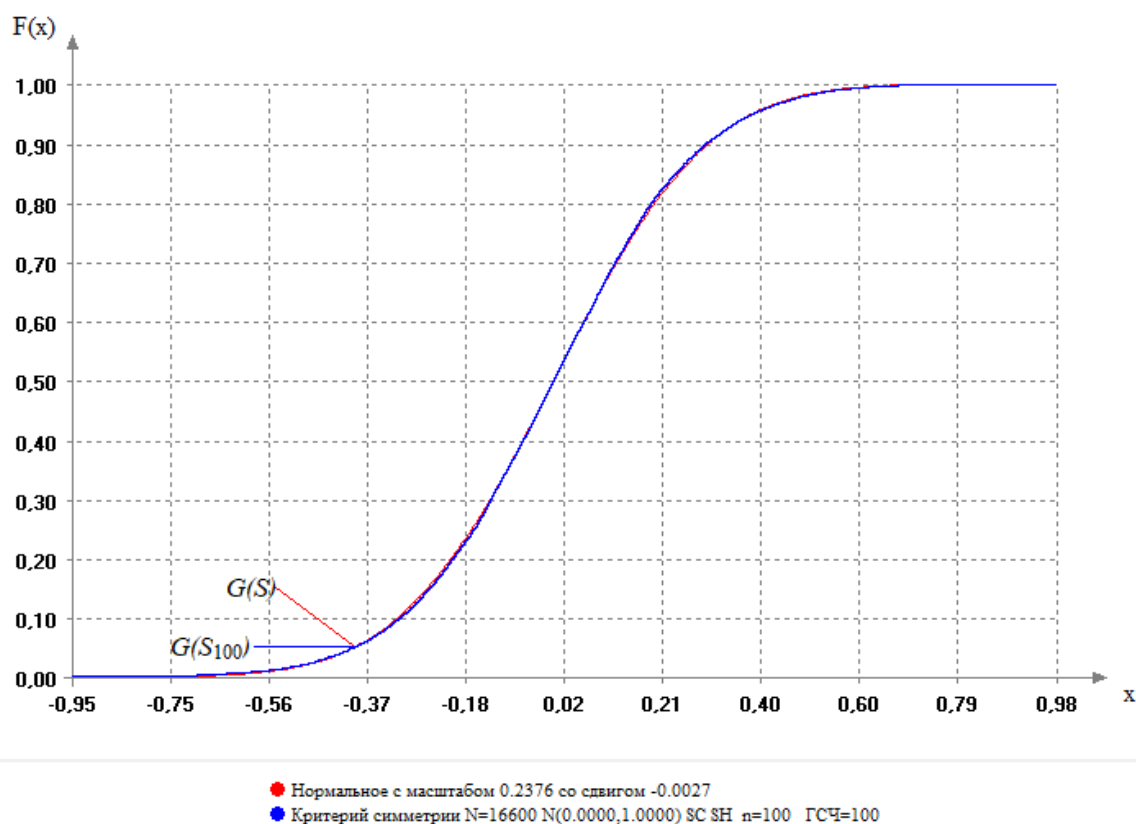
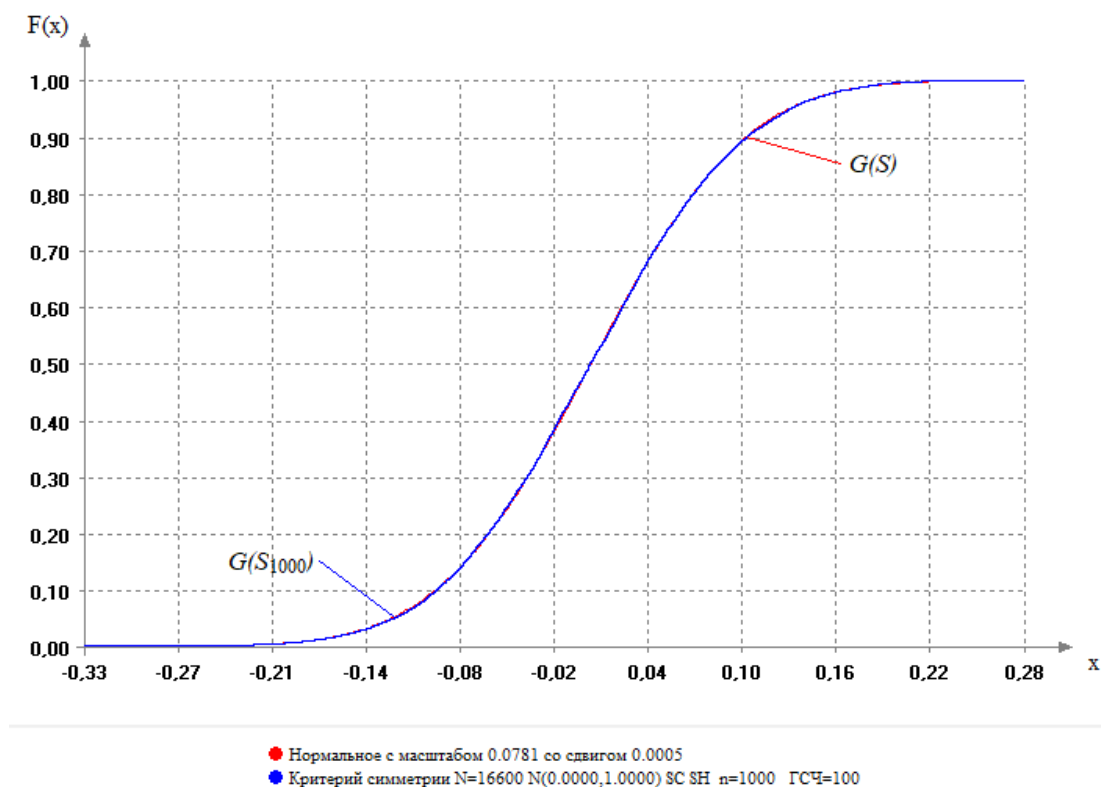
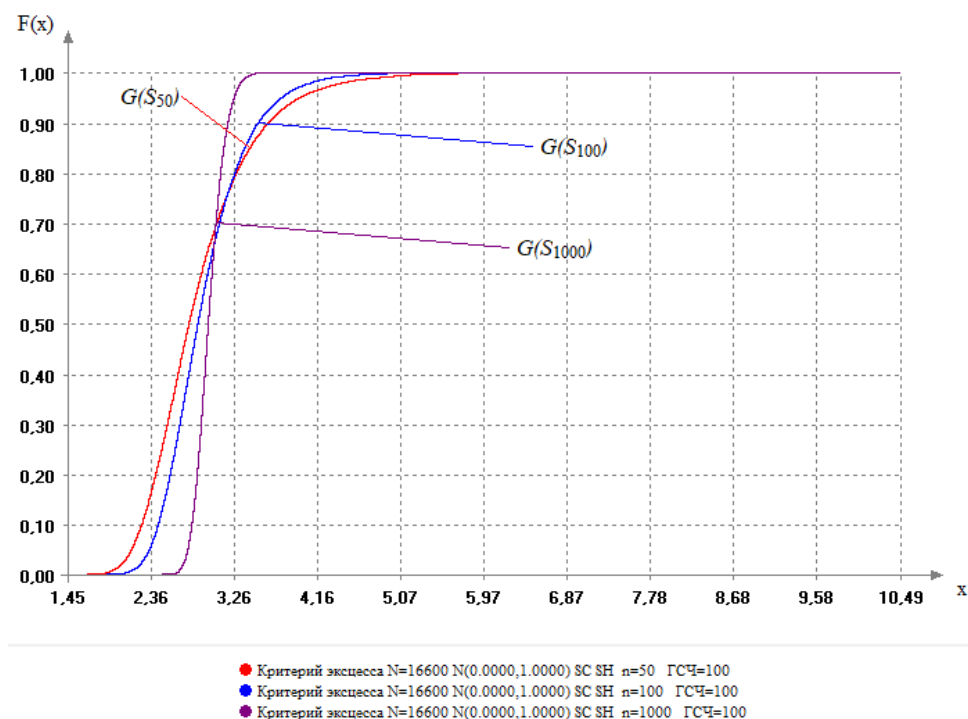


Рисунок 1.3 — Проверка на согласие критерия симметричности при  $n = 100$  с нормальным распределением



Теперь рассмотрим критерий эксцесса для выборок с объёмом  $n = 50, 100, 1000$ . Изобразим их на одном графике:



Опять же ситуация как с критерием симметричности: при увеличении объёма выборки распределения начинают задирается вверх.

Рассмотрим теперь согласие данных статистик с нормальным законом распределения.

$n = 50$	$n = 100$	$n = 1000$
$P = 0$ ОТВЕРГАЕТСЯ	$P = 0$ ОТВЕРГАЕТСЯ	$P = 0$ ОТВЕРГАЕТСЯ

На самом деле, то что все гипотезы были отвержены, не удивительно, так как даже при больших объёмах выборки распределения критерия всё ещё будут далеки от нормального.

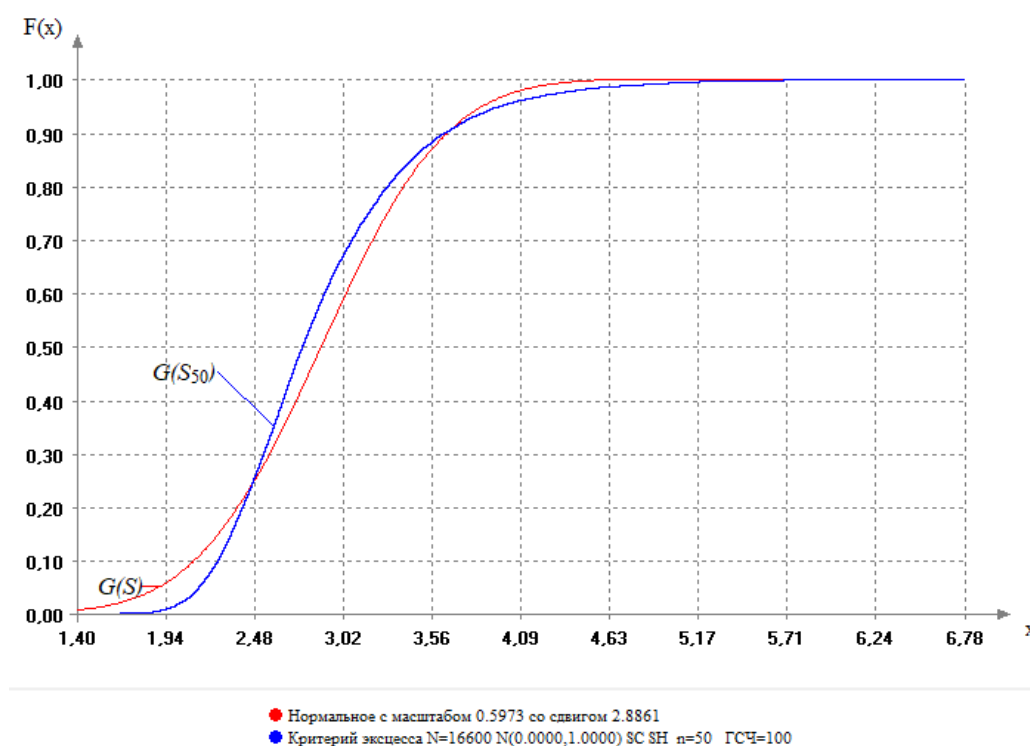


Рисунок 1.6 — Проверка на согласие критерия эксцесса при  $n = 50$  с нормальным распределением

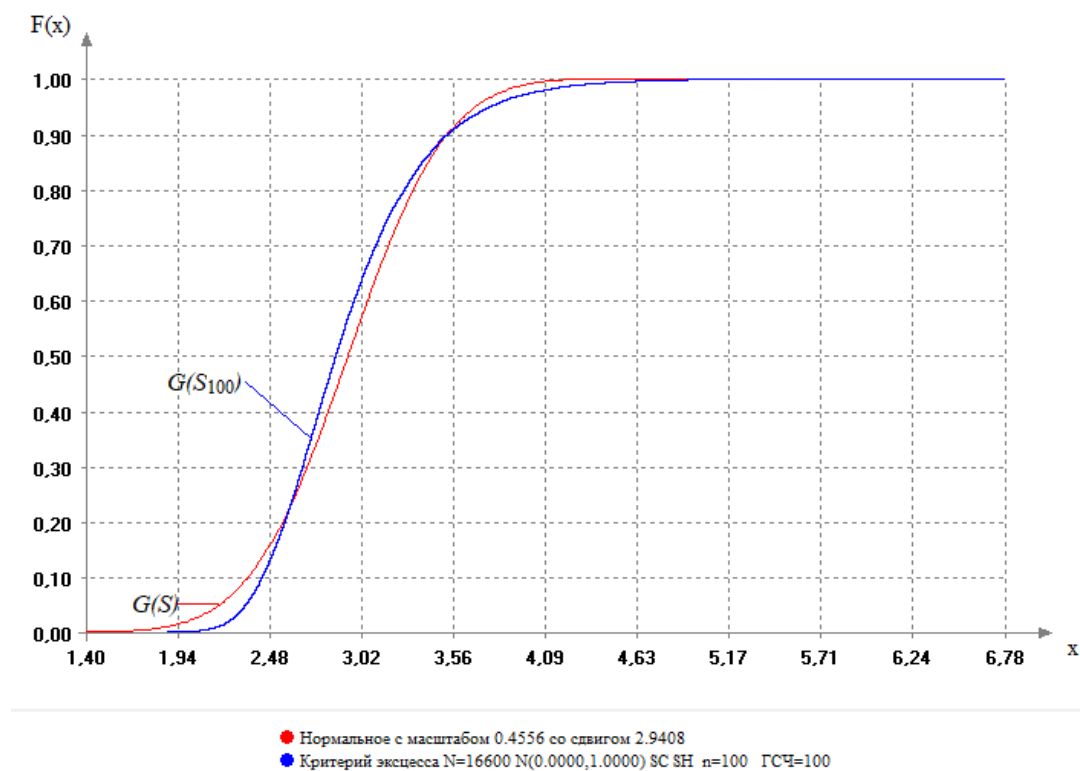


Рисунок 1.7 — Проверка на согласие критерия эксцесса при  $n = 100$  с нормальным распределением

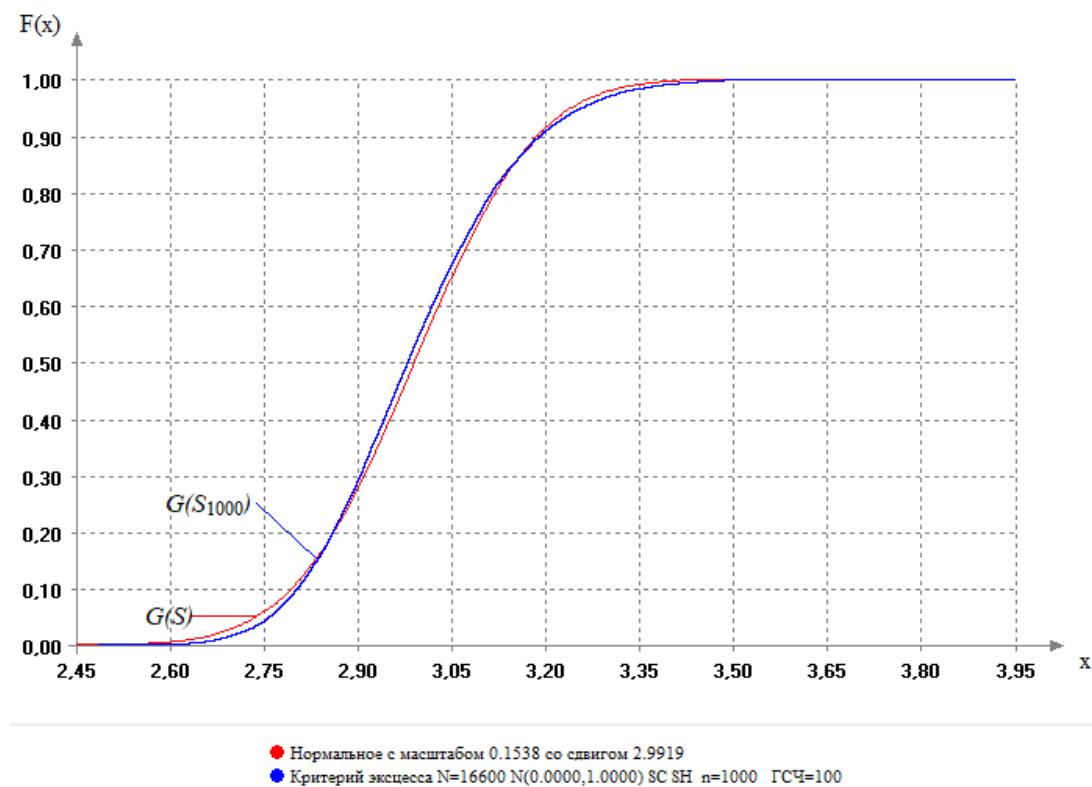


Рисунок 1.8 — Проверка на согласие критерия эксцесса при  $n = 1000$  с нормальным распределением



Теперь рассмотрим критерии симметричности и эксцесса применительно к двустороннему экспоненциальному распределению с формой 5. Объёмы выборок будут те же. После этого посмотрим на согласие полученных распределений с нормальным законом распределения.

### Критерий симметричности:

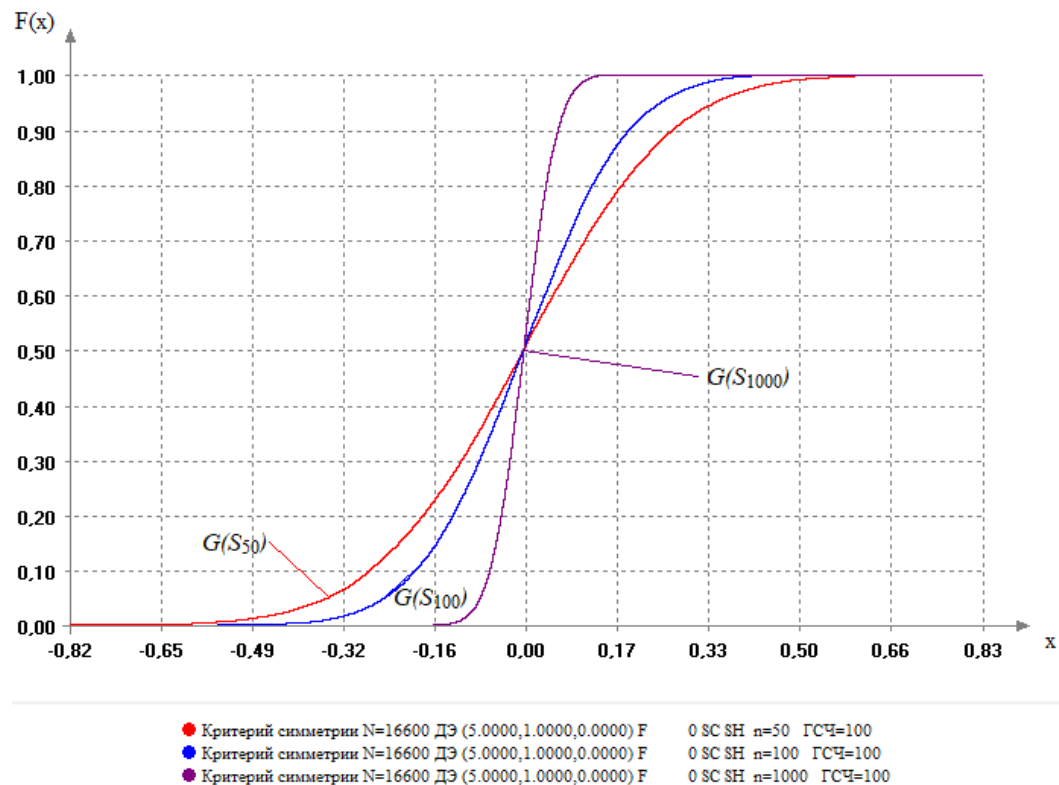


Рисунок 1.9 — Изменение статистики критерия симметричности для двустороннего экспоненциального закона (форма 5) с изменением объёма выборки

$n = 50$	$n = 100$	$n = 1000$
$P = 0.3512747062474696$ <b>НЕ ОТВЕРГАЕТСЯ</b>	$P = 0.4115506980189204$ <b>НЕ ОТВЕРГАЕТСЯ</b>	$P = 0.6213047044939964$ <b>НЕ ОТВЕРГАЕТСЯ</b>

В этот раз результаты получились достаточно интересные. Во всех случаях удалось получить согласие между статистикой критерия симметричности, который был применён к ДЭ (форма 5) и нормальным законом распределения. Такого не произошло даже в предыдущем случае,

когда мы сравнивали два нормальных закона распределения, только на один из них была применена статистика критерия симметричности.

Что же нам даёт этот факт? По сути, не так много, так как получение согласованности по критерию симметричности ещё не подтверждает нормальность нашей эмпирической выборки (критерий необходимый, но недостаточный).

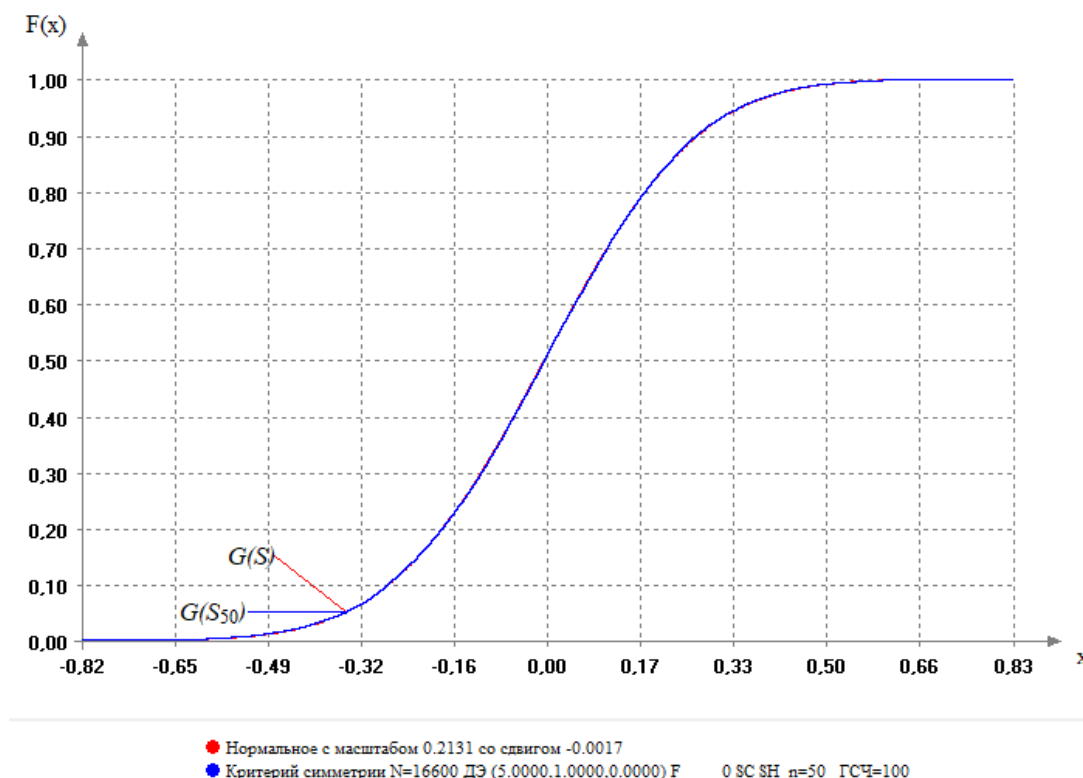


Рисунок 1.10 — Проверка на согласие критерия симметричности для ДЭ (форма 5) при  $n = 50$  с нормальным распределением

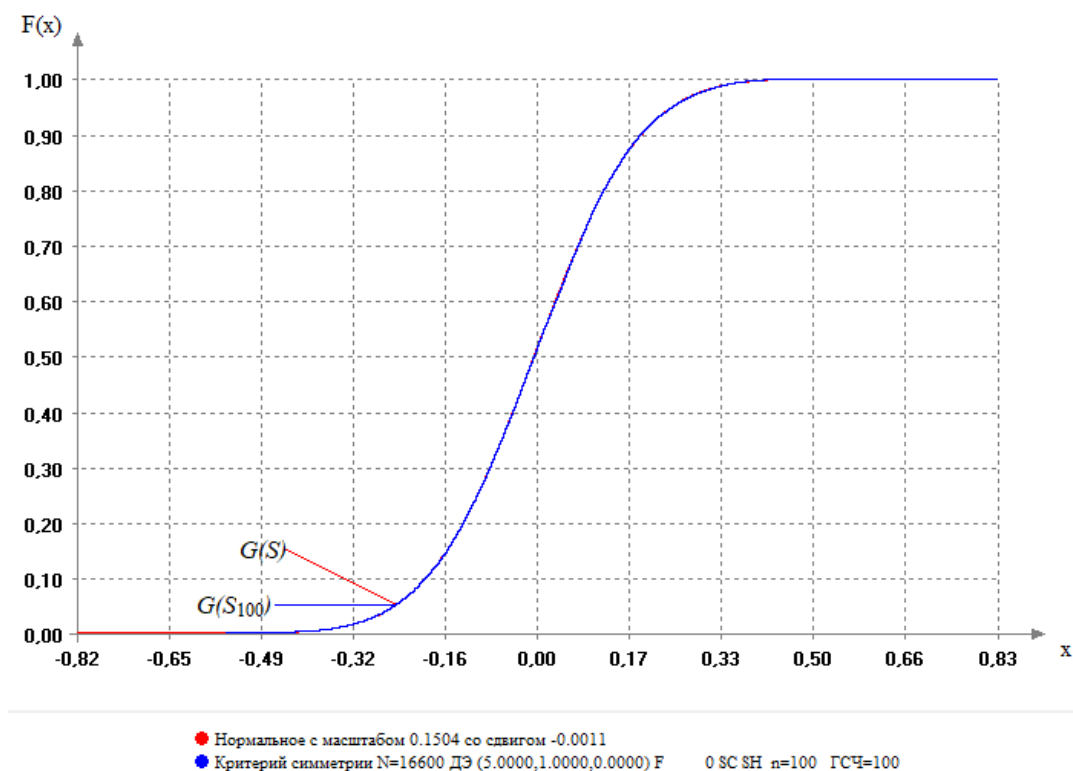


Рисунок 1.11 — Проверка на согласие критерия симметричности для ДЭ (форма 5) при  $n = 100$  с нормальным распределением

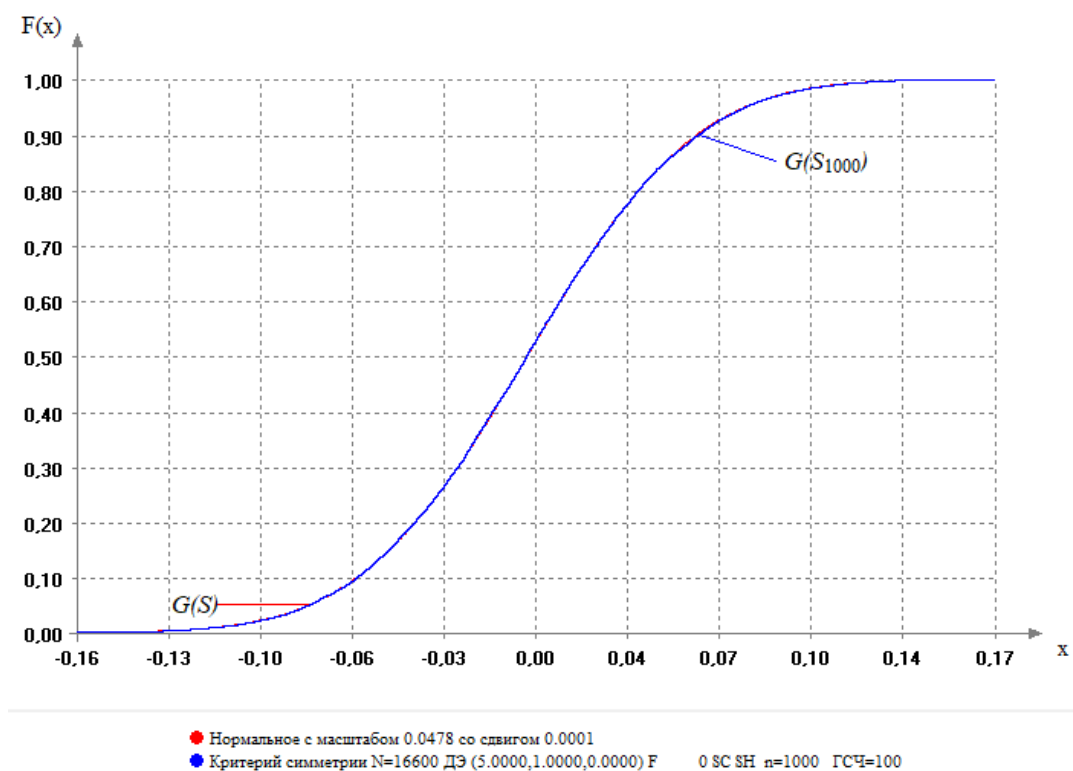


Рисунок 1.12 — Проверка на согласие критерия симметричности для ДЭ (форма 5) при  $n = 1000$  с нормальным распределением

Теперь рассмотрим критерий эксцесса применительно к ДЭ распределению.

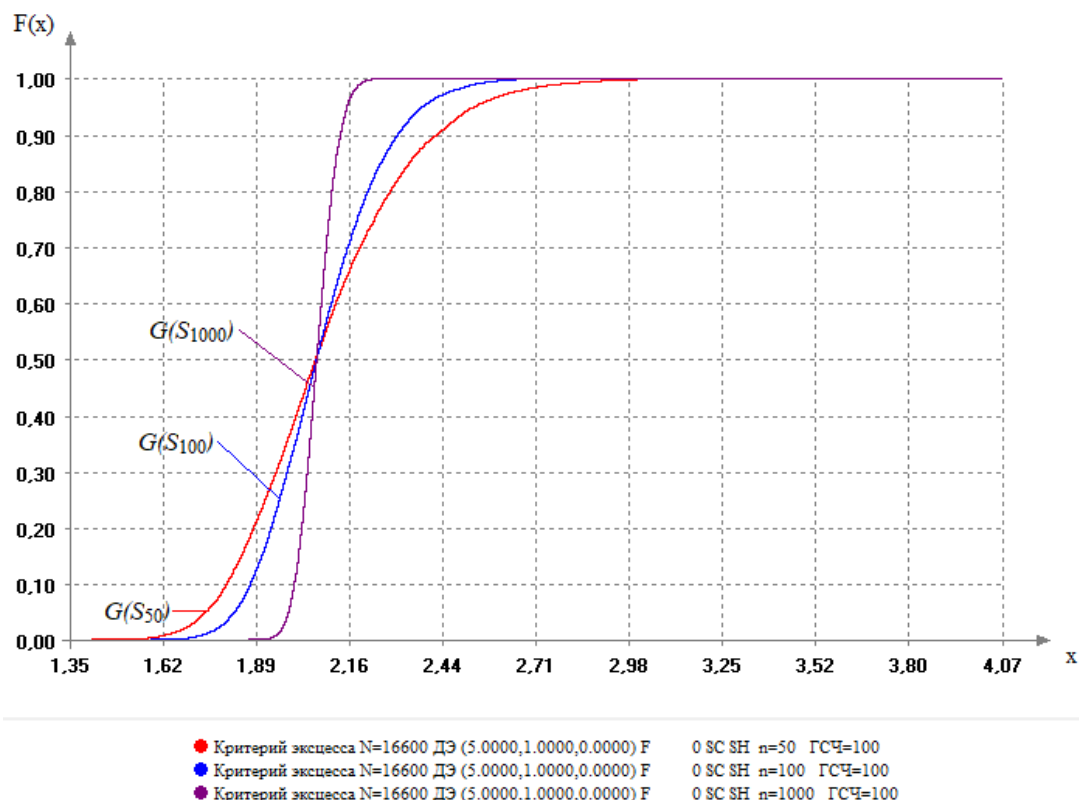


Рисунок 1.13 — Изменение статистики критерия эксцесса для двустороннего экспоненциального закона (форма 5) с изменением объема выборки

$n = 50$	$n = 100$	$n = 1000$
$P = 0$ ОТВЕРГАЕТСЯ	$P = 0$ ОТВЕРГАЕТСЯ	$P = 1.98257136677e - 09$ ОТВЕРГАЕТСЯ

Здесь же ничего принципиально нового мы увидеть не смогли, критерий эксцесса по итогу забраковал все наши распределения из-за недостаточных объемов выборок, возможно необходимы были объёмы выборок ещё больше, дабы удалось получить согласие с нормальным законом, так как при  $n = 1000$  мы хотя бы не получили 0.

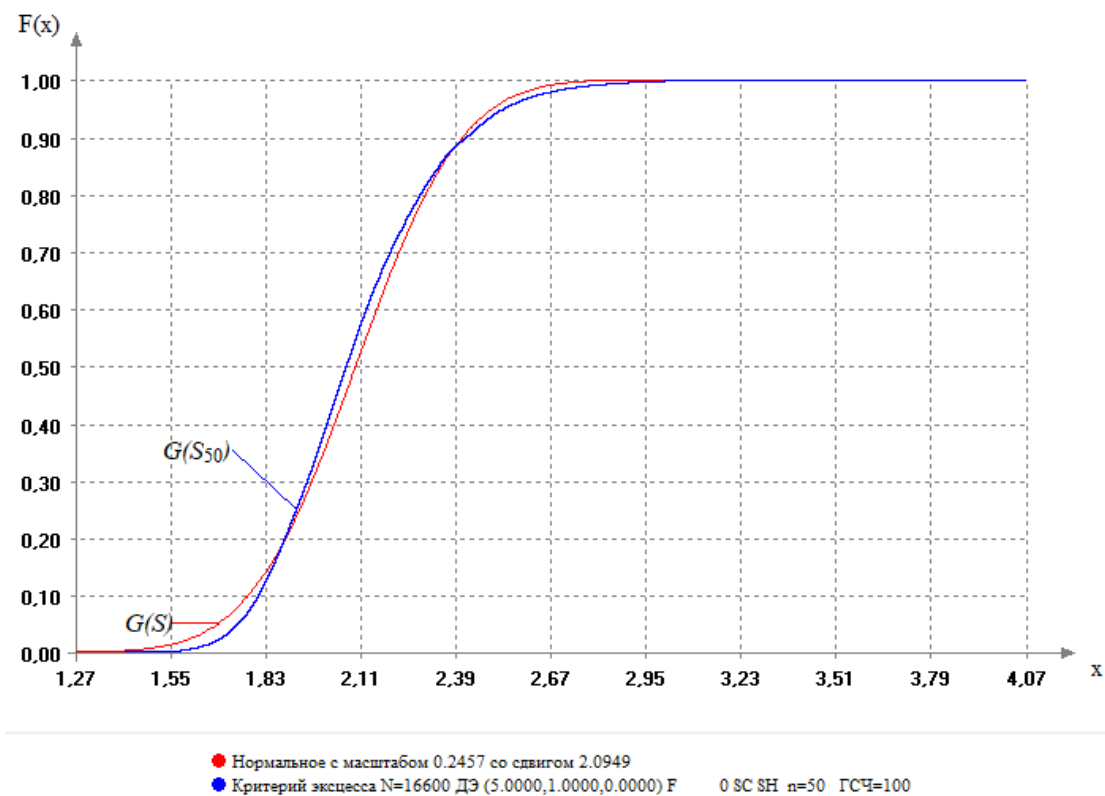


Рисунок 1.14 — Проверка на согласие критерия эксцесса для ДЭ (форма 5) при  $n = 50$  с нормальным распределением

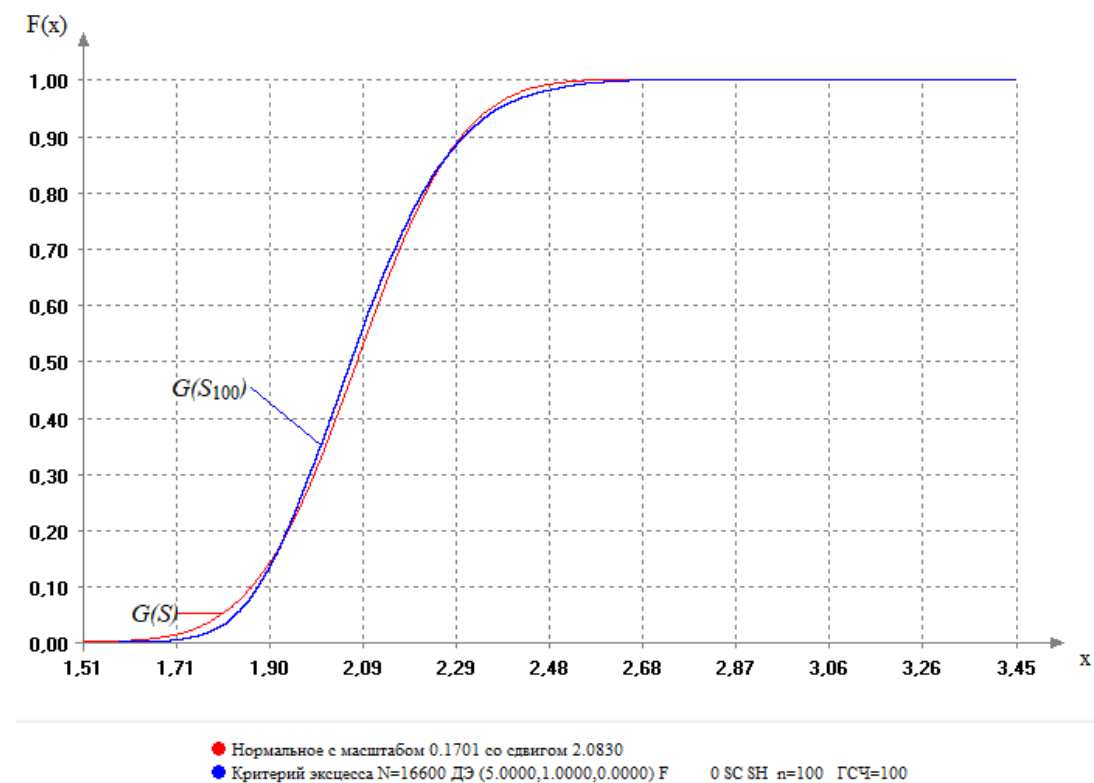
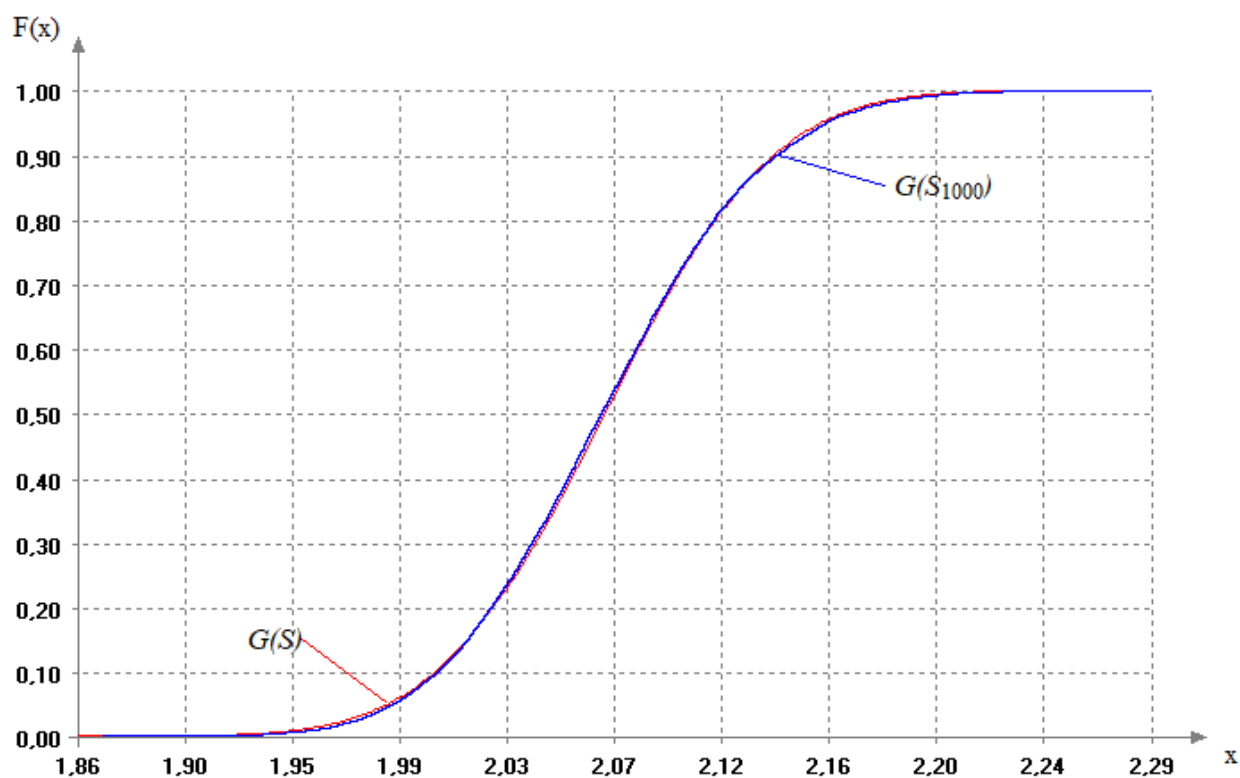


Рисунок 1.15 — Проверка на согласие критерия эксцесса для ДЭ (форма 5) при  $n = 100$  с нормальным распределением



● Нормальное с масштабом 0.0522 со сдвигом 2.0709  
● Критерий эксцесса N=16600 ДЭ (5.0000,1.0000,0.0000) F 0 SC SH n=1000 ГСЧ=100

Рисунок 1.16 — Проверка на согласие критерия эксцесса для ДЭ (форма 5) при  $n = 1000$  с нормальным распределением

**2. Исследовать распределения статистик критерия Шапиро-Уилка при различных объемах выборок. Оценить близость получаемых эмпирических распределений статистик к «теоретическим» по процентным точкам таблиц, соответствующим данному критерию. При некотором объеме выборок ( $n=10$ ) смоделировать распределение статистики критерия при обобщённом нормальном законе (двустороннее экспоненциальное) при параметре формы, равном  $4 \div 7$ . Сравнить с ситуацией, соответствующей справедливой проверяемой гипотезе о нормальном законе. Оценить мощность критерия.**

Распределения статистик критерия Шапиро-Уилка при двустороннем экспоненциальном распределении с объемами выборок  $n = 10, 20, 30, 40$  и  $50$  приведены на рисунке.

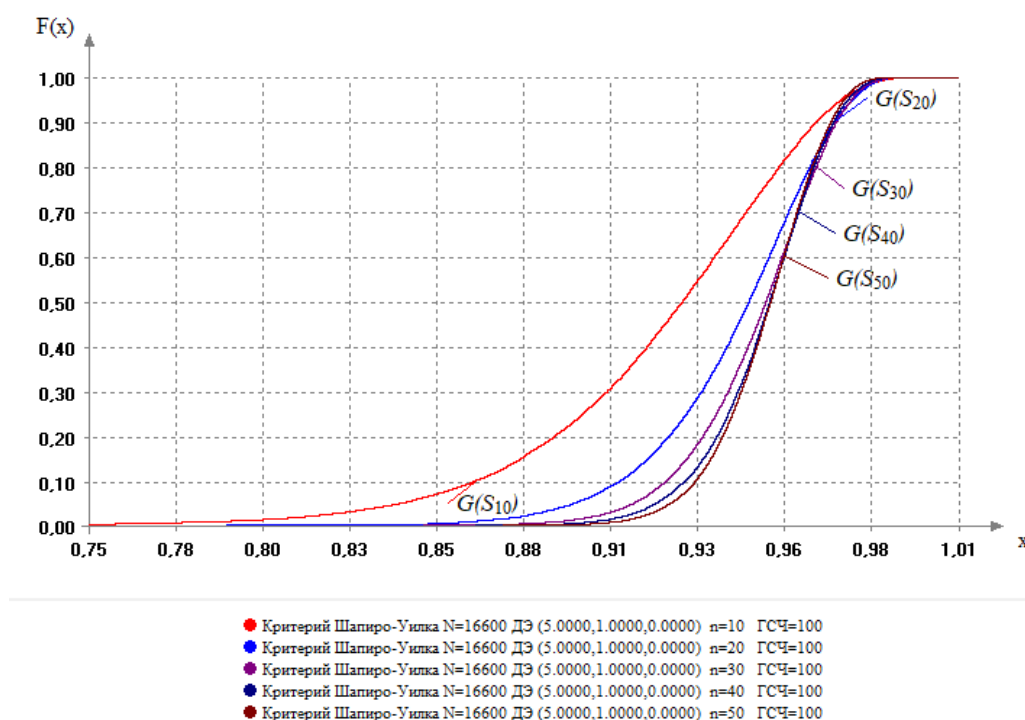


Рисунок 2.1 — Распределения статистик критерия Шапиро-Уилка при двустороннем экспоненциальном распределении с объемами выборок  $n = 10, 20, 30, 40, 50$

Экспериментальные и «теоретические» значения статистик критерия Шапиро-Уилка в процентных точках при разных объемах выборок приведены в таблице.

	$p = 0.01$		$p = 0.05$	
	S, теор	S, эксп	S, теор	S, эксп
$n = 10$	0.781	0.793	0.842	0.844
$n = 20$	0.868	0.867	0.905	0.895
$n = 30$	0.9	0.892	0.927	0.913
$n = 40$	0.919	0.902	0.94	0.92
$n = 50$	0.93	0.91	0.947	0.924

Как можно заметить, ДЭ (форма 5) распределение согласуется во всех случаях с нормальным по статистике критерия Шапиро-Уилка, так как процентные точки практически совпадают при разных уровнях значимости, а также при разных объёмах выборки.

Были смоделированы распределения статистики критерия при обобщенном нормальном распределении с формой 7 и нормальном распределении ( $n = 10$ ). Полученные мощности можно увидеть ниже.

$\alpha$	$1 - \beta$
0.15	0.203976
0.1	0.130843
0.05	0.0599398
0.025	0.0262651
0.01	0.0089759

Основываясь на полученных данных можно сказать, что критерий Шапиро-Уилка для ДЭ и нормального распределений оказывается несмещённым, так как получаемая мощность больше уровня значимости во всех проверенных случаях. Кроме того, мощность получается достаточно низкая, хотя мы проверяем достаточно небольшую выборку, а значит, данный критерий плохо работает при  $n = 10$ , что согласуется с теоретическими данными.



### 3. Аналогично предыдущему пункту исследовать распределения статистики критерия Эппса-Пулли.

Распределения статистик критерия Эппса-Пулли при двустороннем экспоненциальном распределении с объемами выборок  $n = 10, 20, 30, 40, 50, 200$  приведены на рисунке.

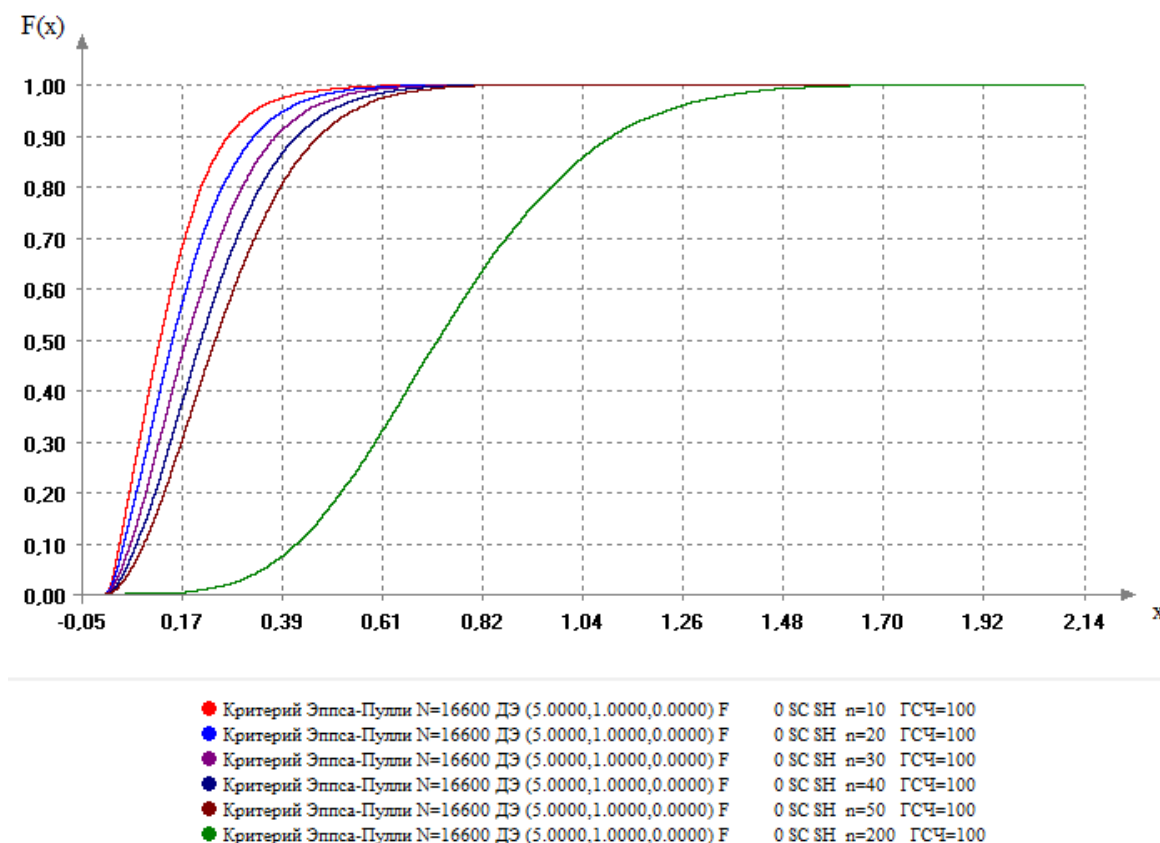


Рисунок 3.1 — распределение статистики Эппса-Пулли применимо к ДЭ (форма 5) распределению при объемах выборки  $n = 10, 20, 30, 40, 50, 200$

Как можно заметить, выборка, объемом в 200 сильно отличается от других и это не удивительно, так как для этого критерия такой объем выборки является предельным (т.к. он применяется при  $8 \leq n \leq 200$ ).

Теперь рассмотрим критические точки при разных уровнях значимости и разных объемах выборки.

	$p = 1 - \alpha$									
	0.85		0.9		0.95		0.975		0.99	
	$S_{\text{теор}}$	$S_{\text{эксп}}$	$S_{\text{теор}}$	$S_{\text{эксп}}$	$S_{\text{теор}}$	$S_{\text{эксп}}$	$S_{\text{теор}}$	$S_{\text{эксп}}$	$S_{\text{теор}}$	$S_{\text{эксп}}$
$n = 10$	0.233	0.2372	0.277	0.2713	0.354	0.3307	0.435	0.395319	0.544	0.493266
$n = 20$	0.240	0.2891	0.287	0.3285	0.368	0.3958	0.452	0.461207	0.565	0.544092
$n = 30$	0.241	0.3331	0.289	0.3758	0.372	0.4439	0.457	0.51503	0.572	0.595346
$n = 40$	0.241	0.3761	0.289	0.4225	0.373	0.4957	0.458	0.569699	0.576	0.65409
$n = 50$	0.241	0.4214	0.290	0.4712	0.374	0.5492	0.460	0.614892	0.577	0.710609
$n = 200$	0.241	1.0355	0.291	1.1120	0.377	1.2347	0.466	1.345	0.585	1.4622

Можно сказать, что наш критерий, базирующийся на функции распределения ДЭ (форма 5) хоть как-то согласуется с нормальным законом распределения лишь при  $n = 10, 20$ . В других же случаях закон распределения критерия должен быть отклонён.

Теперь рассмотрим ДЭ (форма 7) и нормальный закон распределения применимо к критерию Эппса-Пулли.

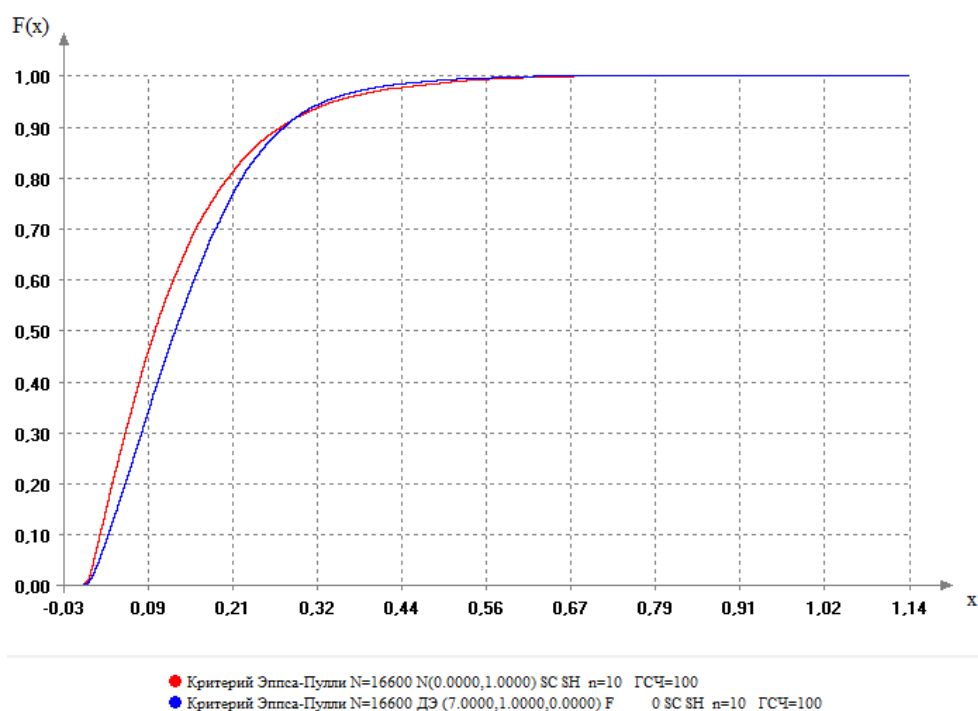


Рисунок 3.2 — сравнение ДЭ (форма 7) и нормального закона распределения, применимо к критерию Эппса-Пулли

$\alpha$	$1 - \beta$
0.15	0,235663
0.1	0,169096
0.05	0,0955422
0.025	0,0529518
0.01	0,0218675

Проблема у данного критерия точно такая же, что и у Шапиро-Уилка, а именно: при малых значениях выборки критерий плохо отличает конкурирующие гипотезы  $H_0$  и  $H_1$ .

**4. Исследовать распределения статистики  $z_1$ . Проверить близость эмпирических распределений статистики стандартному нормальному закону.**

Смоделируем для  $z_1$  ДЭ (форма 5) распределение с объёмами выборки  $n = 10, 20, 30, 40, 50, 100$ .

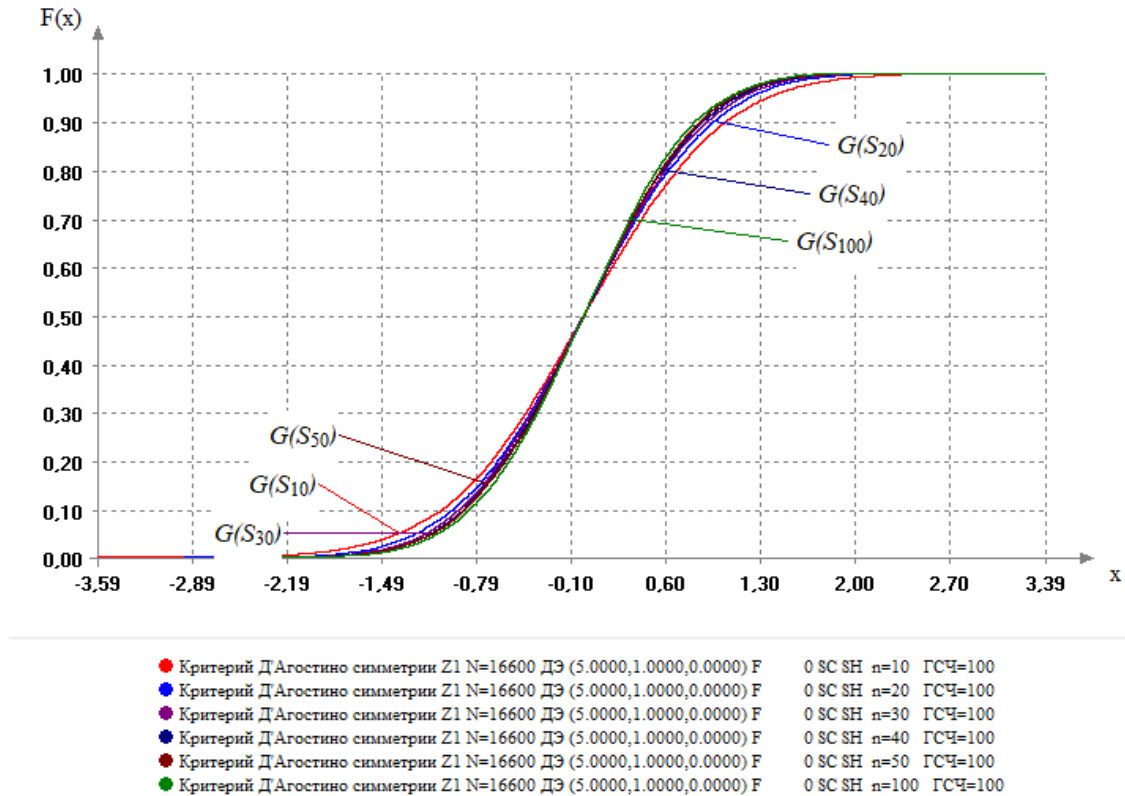


Рисунок 4.1 — распределение статистики  $z_1$  применимо к ДЭ (форма 5) распределению при объёмах выборки  $n = 10, 20, 30, 40, 50, 100$

Здесь картина немного отличается от предыдущих критериев нормальности. Распределения статистики пересекаются друг с другом в примерно одной области, в прошлые разы же у нас графики просто съезжали вправо при увеличении объёма выборки.

Теперь посмотрим на согласие нашего критерия нормальности, рассчитанного на основе ДЭ (форма 5) с нормальным законом распределения для  $n = 10, 20, 30, 40, 50, 100$ .

Объём	$P = 1 - G(S H_0)$ при $\alpha = 0.1$	Результат
$n = 10$	0.00494473385735215	ОТВЕРГАЕТСЯ
$n = 20$	0.7390609097816183	НЕ ОТВЕРГАЕТСЯ
$n = 30$	0.04020531025833706	ОТВЕРГАЕТСЯ
$n = 40$	0.3202343585464279	НЕ ОТВЕРГАЕТСЯ
$n = 50$	0.1206889876794788	НЕ ОТВЕРГАЕТСЯ
$n = 100$	0.122302392157546	НЕ ОТВЕРГАЕТСЯ

В итоге получается картина, схожая с критерием симметричности. У нас при небольшом изменении объёма выборки критерий D'Agostino может как согласовываться с нормальным законом распределения, так и отвергаться.

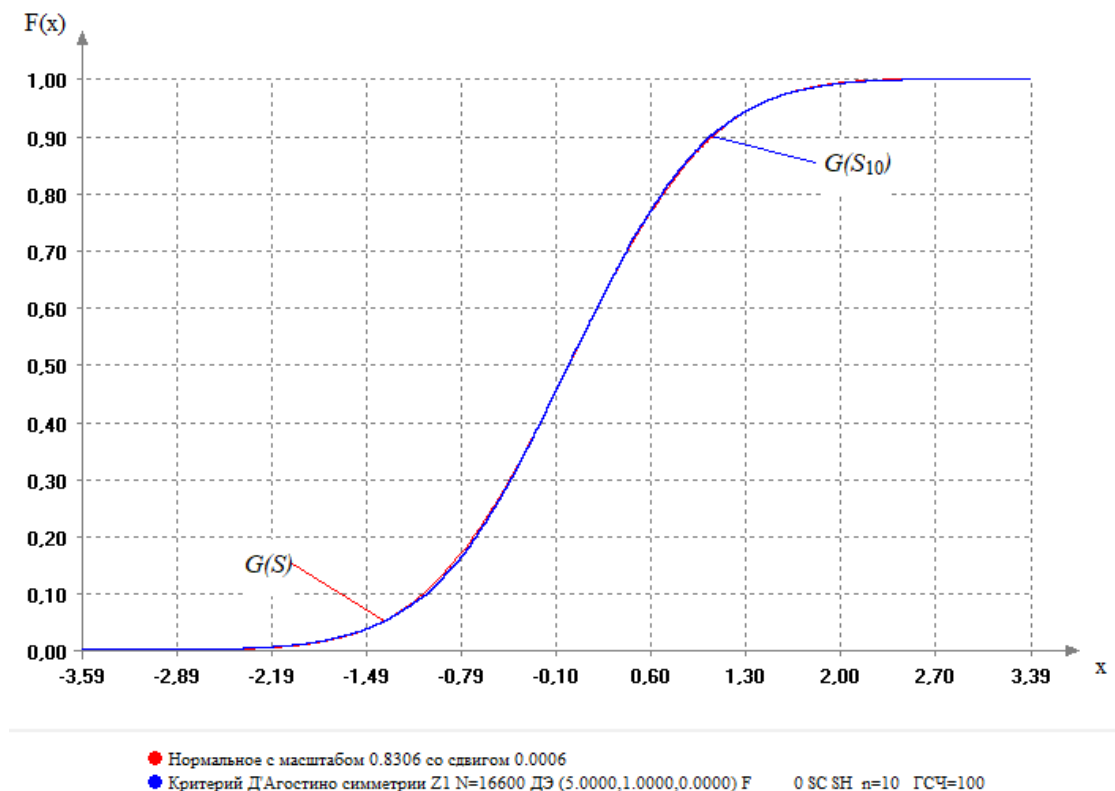


Рисунок 4.2 — Проверка на согласие критерия  $z_1$  для ДЭ (форма 5) при  $n = 10$  с нормальным распределением

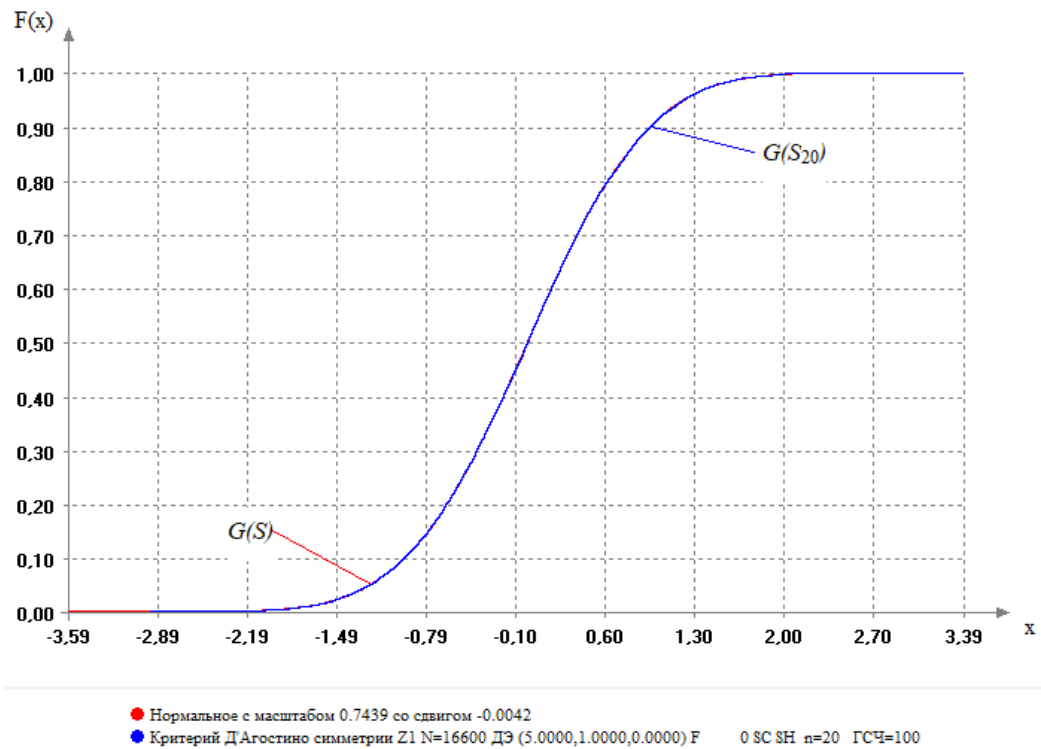


Рисунок 4.3 — Проверка на согласие критерия  $z_1$  для ДЭ (форма 5) при  $n = 20$  с нормальным распределением

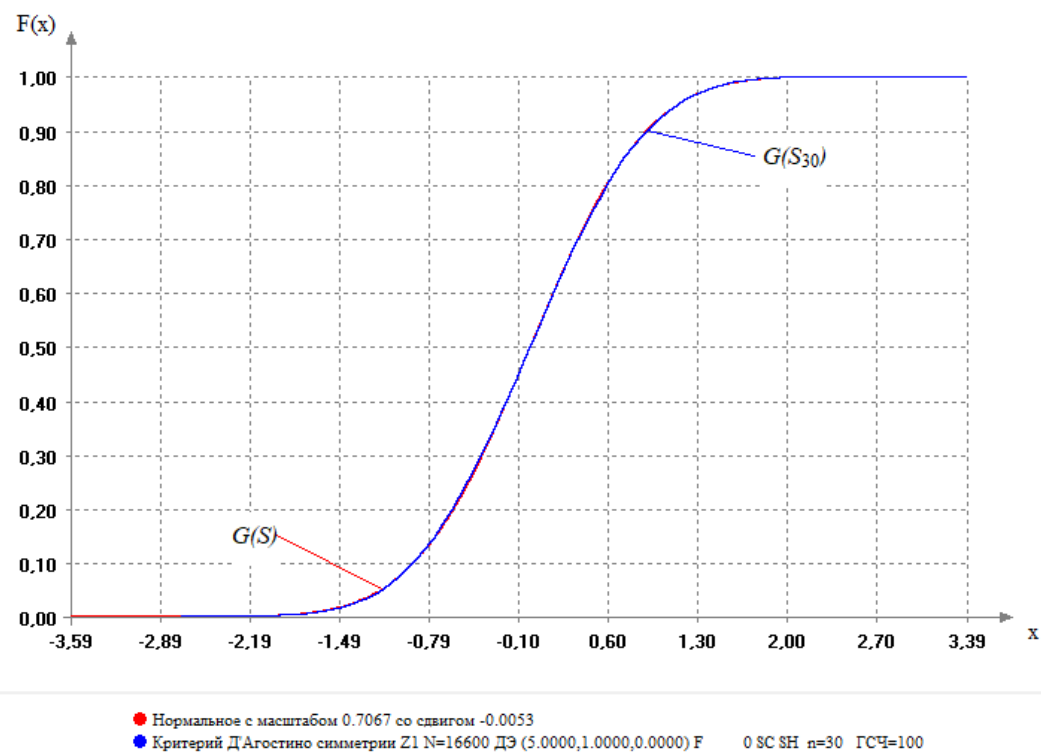


Рисунок 4.3 — Проверка на согласие критерия  $z_1$  для ДЭ (форма 5) при  $n = 30$  с нормальным распределением

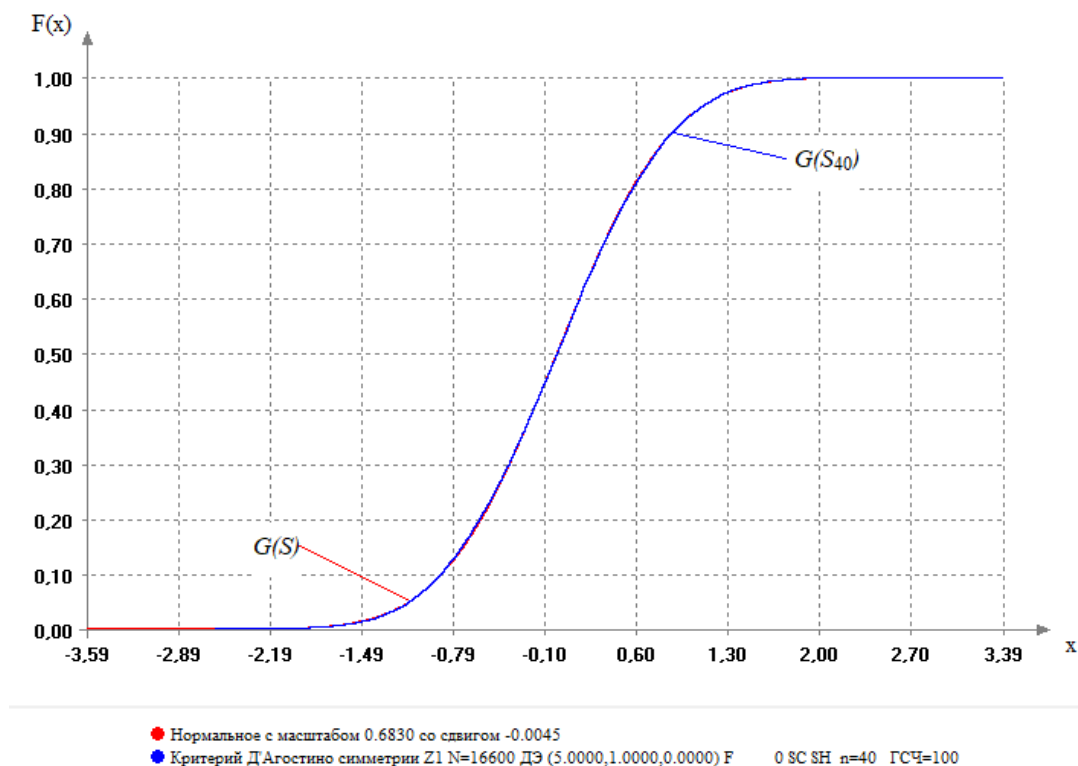


Рисунок 4.4 — Проверка на согласие критерия  $z_1$  для ДЭ (форма 5) при  $n = 40$  с нормальным распределением

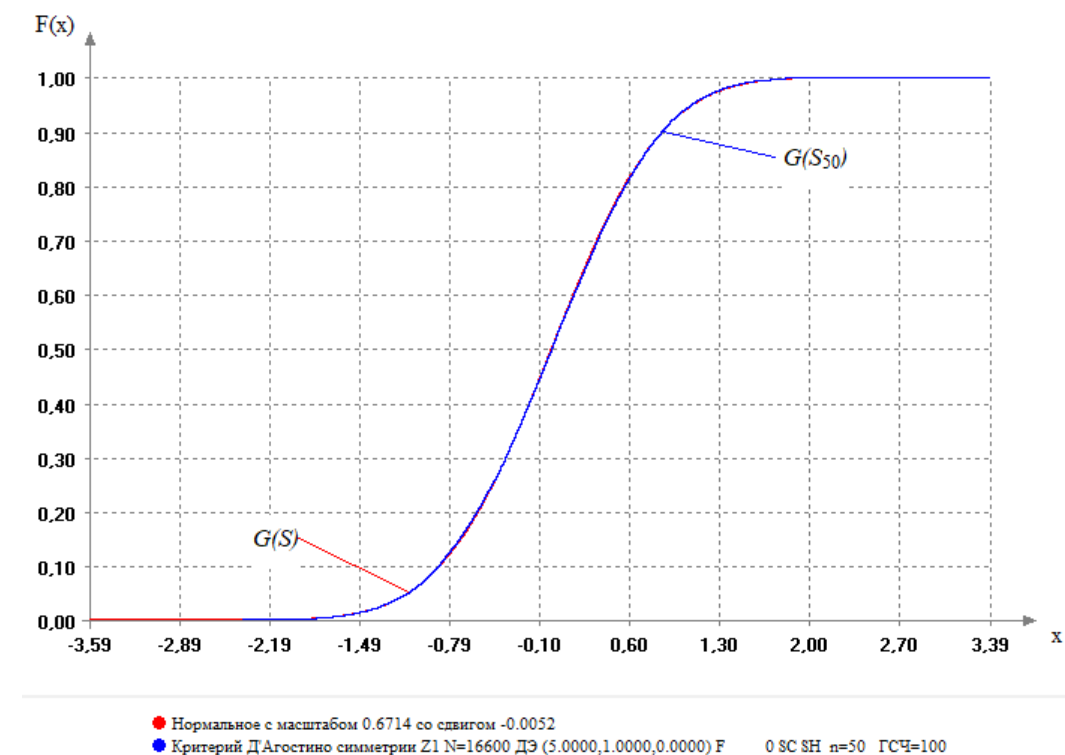


Рисунок 4.5 — Проверка на согласие критерия  $z_1$  для ДЭ (форма 5) при  $n = 50$  с нормальным распределением

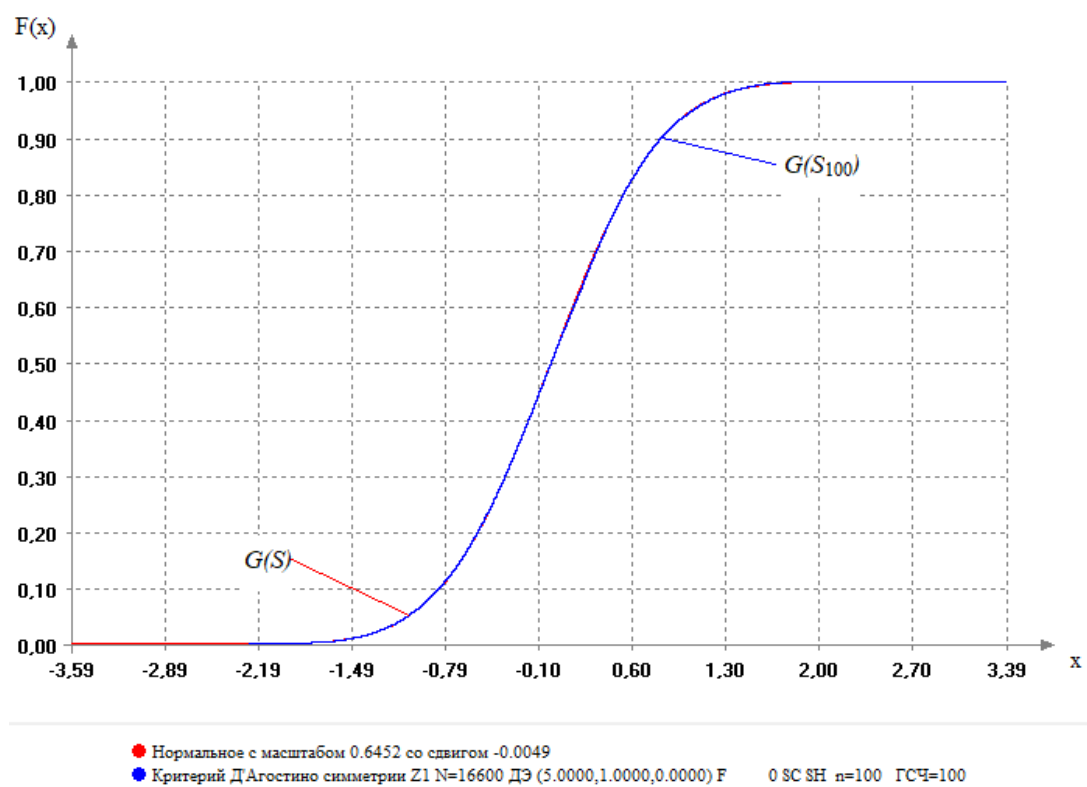


Рисунок 4.6 — Проверка на согласие критерия  $z_1$  для ДЭ (форма 5) при  $n = 100$  с нормальным распределением



**5. Исследовать распределения статистики  $z_2$ . Проверить близость эмпирических распределений статистики стандартному нормальному закону.**

Смоделируем для  $z_2$  ДЭ (форма 5) распределение с объёмами выборки  $n = 10, 20, 30, 40, 50, 100$ .

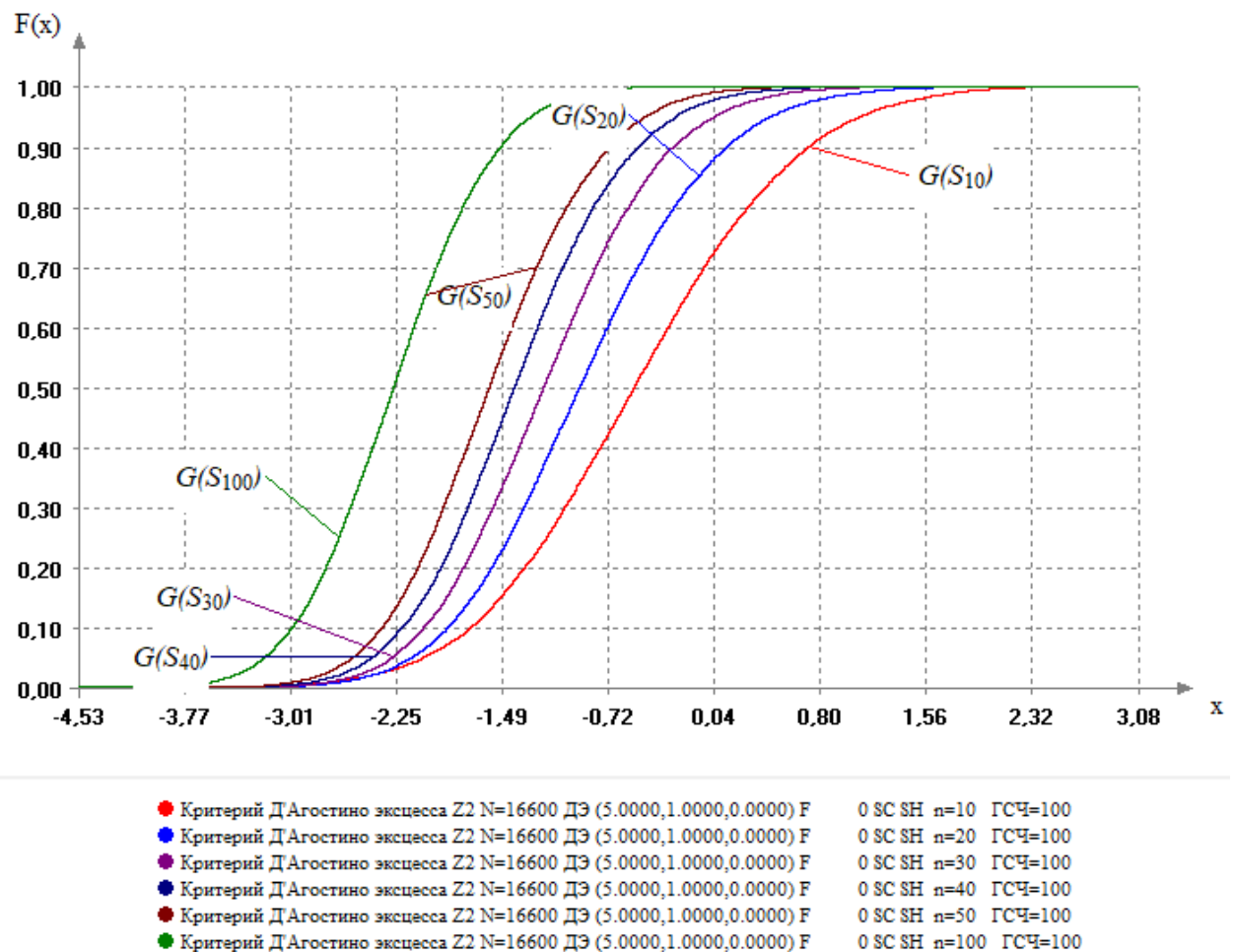


Рисунок 5.1 — распределение статистики  $z_2$  применимо к ДЭ (форма 5) распределению при объёмах выборки  $n = 10, 20, 30, 40, 50, 100$

Здесь ситуация схожа с другими критериями нормальности: при увеличении объёма выборки распределения смещаются вправо.

Теперь посмотрим, как согласуется данный критерий с нормальным законом распределения.

Объём	$P = 1 - G(S H_0)$ при $\alpha = 0.1$	Результат
$n = 10$	3.944990160272027e-07	ОТВЕРГАЕТСЯ
$n = 20$	4.153095680870647e-30	ОТВЕРГАЕТСЯ
$n = 30$	6.948345591822793e-24	ОТВЕРГАЕТСЯ
$n = 40$	1.465430953799735e-24	ОТВЕРГАЕТСЯ
$n = 50$	5.793140233261307e-24	ОТВЕРГАЕТСЯ
$n = 100$	3.255674551405189e-11	ОТВЕРГАЕТСЯ
$n = 1000$	0.4095827437345858	НЕ ОТВЕРГАЕТСЯ

Несмотря на то, что у нас визуально данный критерий сходится с нормальным законом распределения, но в итоге всё равно гипотеза о нормальности отвергается в большинстве случаев. Но зато при увеличении объёма выборки значение  $P$  увеличивается вплоть до согласия с нормальным распределением.

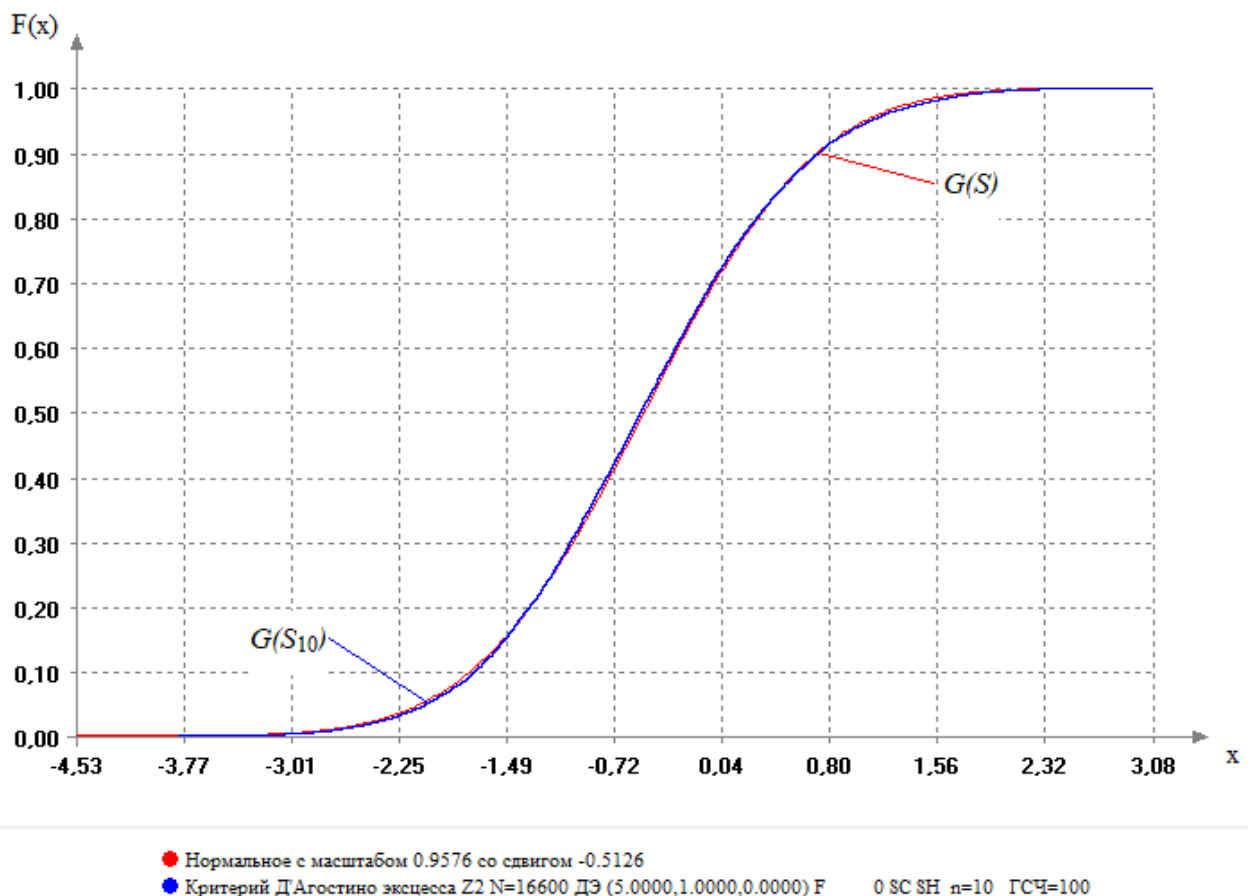


Рисунок 5.2 — Проверка на согласие критерия  $z_2$  для ДЭ (форма 5) при  $n = 10$  с нормальным распределением

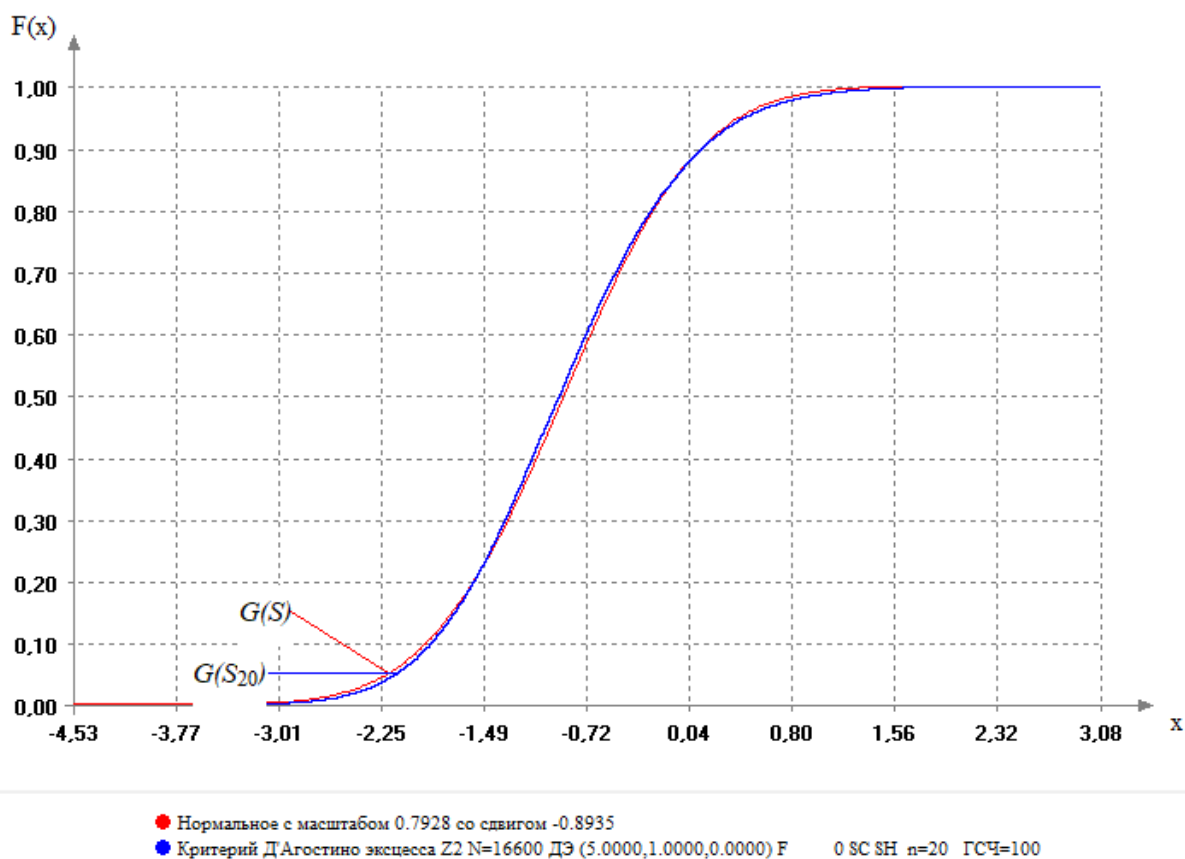


Рисунок 5.3 — Проверка на согласие критерия  $z_2$  для ДЭ (форма 5) при  $n = 20$  с нормальным распределением

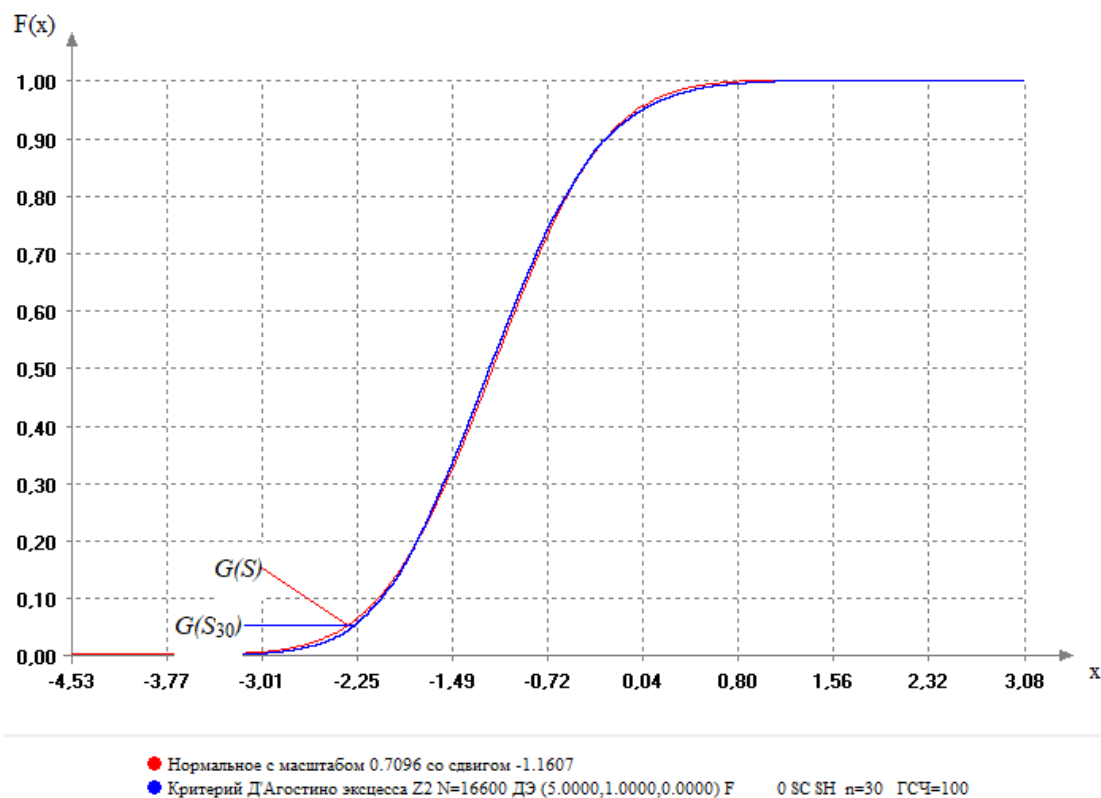


Рисунок 5.4 — Проверка на согласие критерия  $z_2$  для ДЭ (форма 5) при  $n = 30$  с нормальным распределением

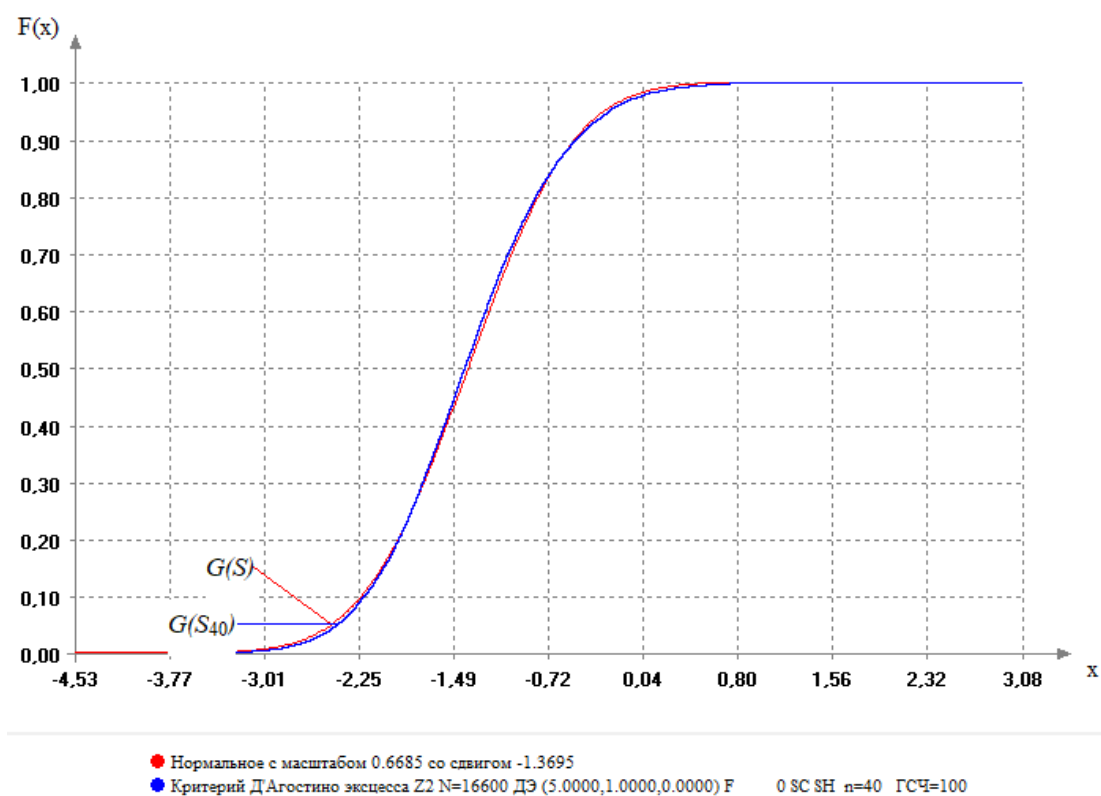


Рисунок 5.5 — Проверка на согласие критерия  $z_2$  для ДЭ (форма 5) при  $n = 40$  с нормальным распределением

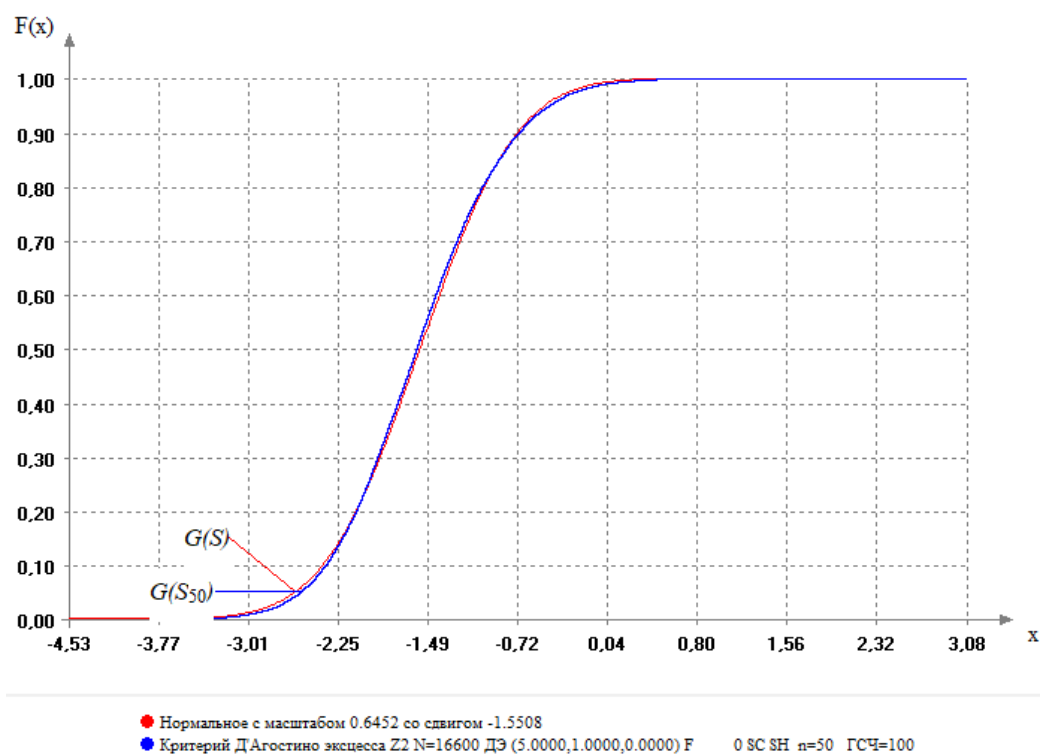


Рисунок 5.6 — Проверка на согласие критерия  $z_2$  для ДЭ (форма 5) при  $n = 50$  с нормальным распределением

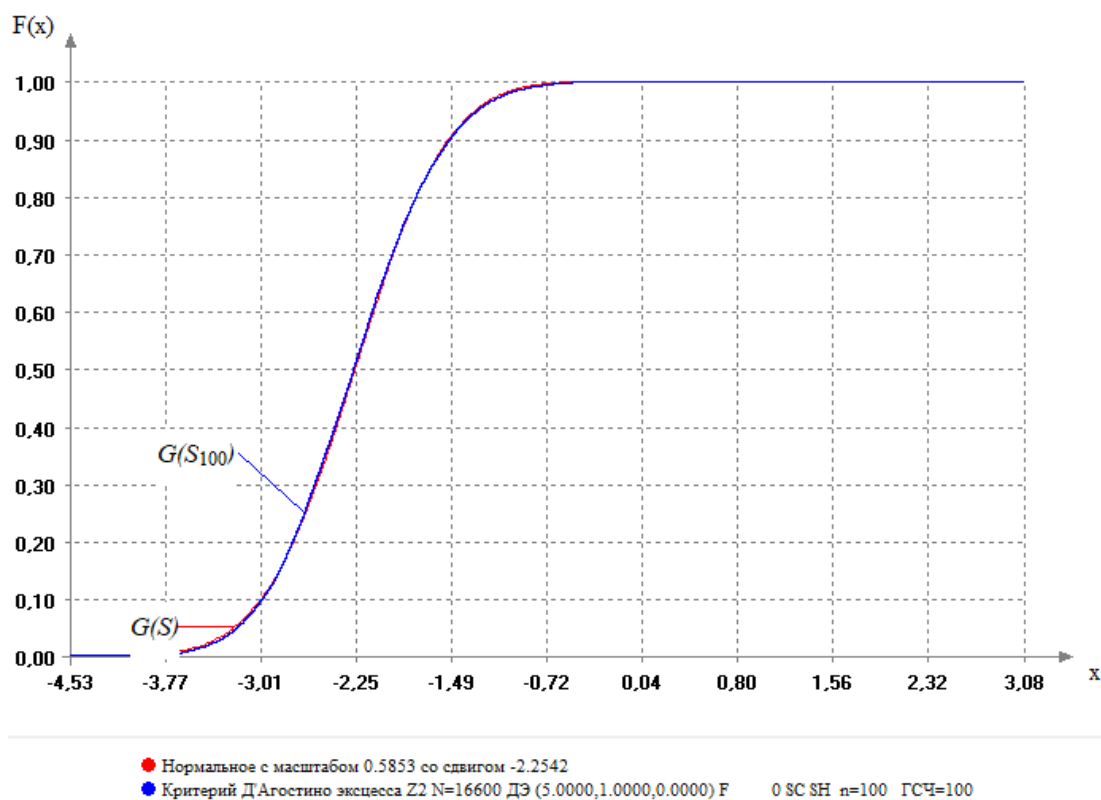


Рисунок 5.7 — Проверка на согласие критерия  $z_2$  для ДЭ (форма 5) при  $n = 100$  с нормальным распределением

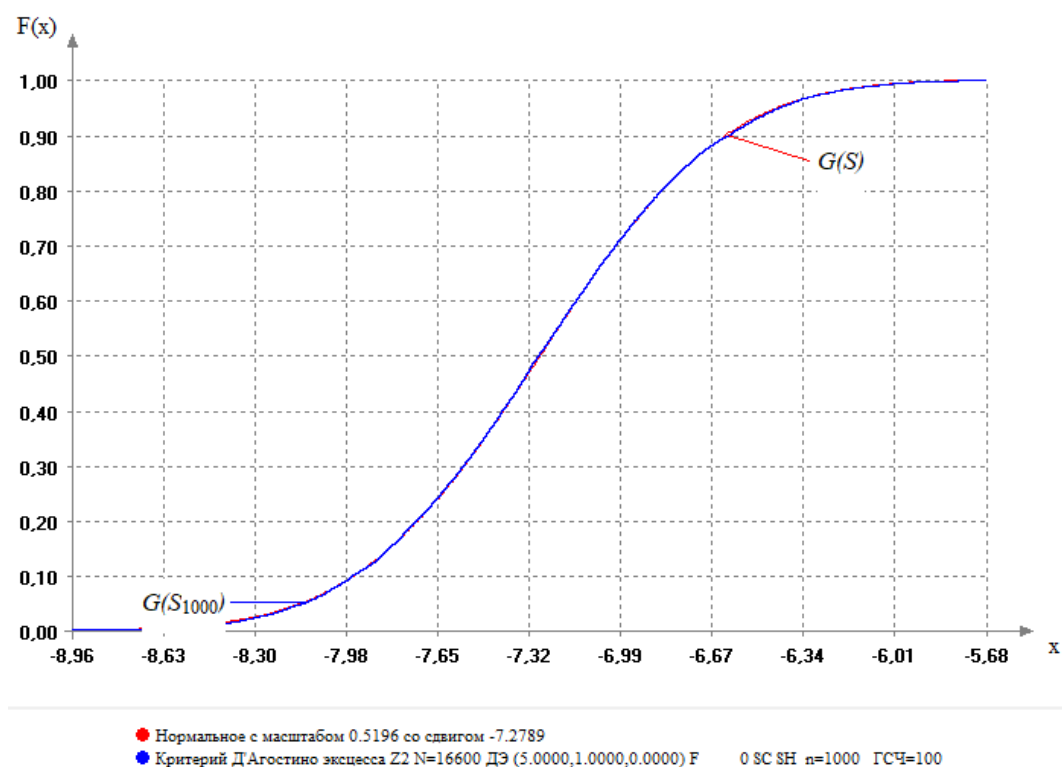


Рисунок 5.8 — Проверка на согласие критерия  $z_2$  для ДЭ (форма 5) при  $n = 1000$  с нормальным распределением

**6. Оценить мощность критериев со статистиками (7.3), (7.4), (7.5), (7.6) относительно заданной альтернативы.**

**Для  $n = 10$**

$\alpha$	Шапиро-Уилк	Эппса-Пулли	$z_1$	$z_2$
0.15	0.203976	0,235663	0.197831	0.245663
0.1	0.130843	0,169096	0.149157	0.179819
0.05	0.0599398	0,0955422	0.0866867	0.0981928
0.025	0.0262651	0,0529518	0.0496988	0.0557831
0.01	0.0089759	0,0218675	0.0268675	0.0255422

### **Вывод:**

Выяснилось, что наиболее эффективным критерием нормальности при исследовании объёмов выборок  $n = 10$  оказался  $z_2$ , далее за ним стоит по эффективности критерий Эппса-Пулли, но он всё равно достаточно плохо работает в интервале  $n = 10..20$  и практически не различает гипотезы  $H_0$  и  $H_1$ . После него по эффективности идёт  $z_1$ , а за ним располагается критерий Шапиро-Уилка. Проблема Шапиро-Уилка заключается в плохой работе с выборками объёмом  $n = 10..20$ , он также не различает гипотезы  $H_0$  и  $H_1$ . Проблема  $z_1$  состоит исключительно в рассмотрении одного критерия – симметричности, из-за чего он работает не всегда однозначно: например, как мы видели выше, если к согласующийся выборке добавить 10 эмпирических наблюдений, данный критерий уже может быть отвергнут.

Градация по мощностям следующая:

- 1)  $z_2$
- 2) Эппса-Пулли
- 3)  $z_1$
- 4) Шапиро-Уилка