XXV НАУЧНО-ПРАКТИЧЕСКАЯ КОНФЕРЕНЦИЯ
«КОМПЛЕКСНАЯ ЗАЩИТА ИНФОРМАЦИИ»,
Московская область, 15 – 17 сентября 2020 года

СТАТИСТИЧЕСКОЕ ТЕСТИРОВАНИЕ ПОСЛЕДОВАТЕЛЬНОСТЕЙ С ПРИМЕНЕНИЕМ ЗАКОНА ПОВТОРНОГО ЛОГАРИФМА

НИИ прикладных проблем
математики и информатики БГУ

А.И.Трубей, М.В.Мальцев,
В.Ю.Палуха, И.К.Пирштук

ВВЕДЕНИЕ

Для оценки качества генераторов, применяемых в целях защиты данных, часто используется набор тестов (батарея) NIST SP800-22, чтобы обнаружить отклонения двоичной последовательности от модели независимых симметричных испытаний Бернулли.

Однако батарея имеет существенные недостатки, связанные с ошибками 2 рода. Генератор, который, в основном, генерирует случайные последовательности, с вероятностью β будет также генерировать смещенные на некоторую величину Δ от равновероятного распределения последовательности (например, последовательности, состоящие, в основном, из нулей или единиц). При этом генератор будет оцениваться как «хороший» тестами NIST SP800-22, хотя выходные последовательности несложно отличить от равномерного распределения.

ВВЕДЕНИЕ

Кроме того, NIST SP800-22 не охватывает некоторые основополагающие законы случайности. Существуют две фундаментальные предельные теоремы о случайных двоичных последовательностях – это центральная предельная теорема и закон повторного логарифма (ЗПЛ). Несколько тестов в NIST SP800-22 включают центральную предельную теорему, в то время как ни один тест не охватывает закон повторного логарифма.

В докладе приводятся преимущества тестирования случайных последовательностей, основанные на статистических расстояниях и законе повторного логарифма. Описывается методика принятия решений о качестве последовательностей на основе статистического расстояния с использованием статистики хи-квадрат согласия.

1. Тестирование генераторов псевдослучайных последовательностей

Линейный конгруэнтный генератор (ЛКГ) определяется рекуррентным соотношением

$X_{n+1} = aX_n + c \ (\mathrm{mod} m)$, где m – модуль, a, с $< m$.

Для любого начального значения X_0 псевдослучайная последовательность имеет вид $X_0, X_1, \ldots, X_i \ldots$, где X_i – двоичное представление целого числа X_i.

ЛКГ были включены в различные языки программирования, например, в C и C++. Функции drand48(), rand48(), mrand48(), и rand48() генерируют равномерно распределенные случайные числа по формуле: $X_{n+1} = 0\text{x}343\text{FD} \cdot X_n + 0\text{x}269\text{EC3} (mod 2^{32})$

(выбираются 16–30 биты)

1. Тестирование генераторов псевдослучайных последовательностей. Батарея NIST SP800-22

Проведены две сессии тестирования с количеством выборок 1000 и 5000, а также следующими параметрами тестирования:

- объем выборки – 10^6 бит (125 000 байт);
- уровень значимости на первом этапе – 0,01;
- уровень значимости на втором этапе – 0,0001;

В результате было установлено:

- при числе выборок 1000 – тестирование пройдено успешно;
- при числе выборок 5000 – тестирование пройдено успешно.

Для сравнения было проведено также тестирование последовательностей, выработанных физическим генератором на основе шумового диода «Ключ-04», которое также не выявило отклонений от нулевой гипотезы.

2. Закон повторного логарифма для равновероятной последовательности

Пусть независимые случайные величины $X = (x_1, \dots, x_n)$ одинаково распределены и принимают два значения 0 и 1 с вероятностью $^1/_2$. Тогда $S_n = \sum_{i=0}^{n-1} x_i$ - число успехов в схеме Бернулли с вероятностью успеха $^1/_2$. Э. Борель доказал, что при $n \to \infty$ с вероятностью 1 $\frac{S_n}{n} \to \frac{1}{2}$.

Впоследствии (1914) Г. Харди и Дж. Литлвуд показали, что почти наверное

$$\limsup_{n\to\infty} \frac{\left|S_n - \frac{n}{2}\right|}{\sqrt{n \ln n}} < \frac{1}{\sqrt{2}}.$$

Затем А. Я. Хинчин (1924) доказал более сильный результат, называемый законом повторного логарифма :

$$P\left(\limsup_{n\to\infty} \frac{\left|S_n - \frac{n}{2}\right|}{\sqrt{n \ln \ln n}} = \frac{1}{\sqrt{2}}\right) = 1$$

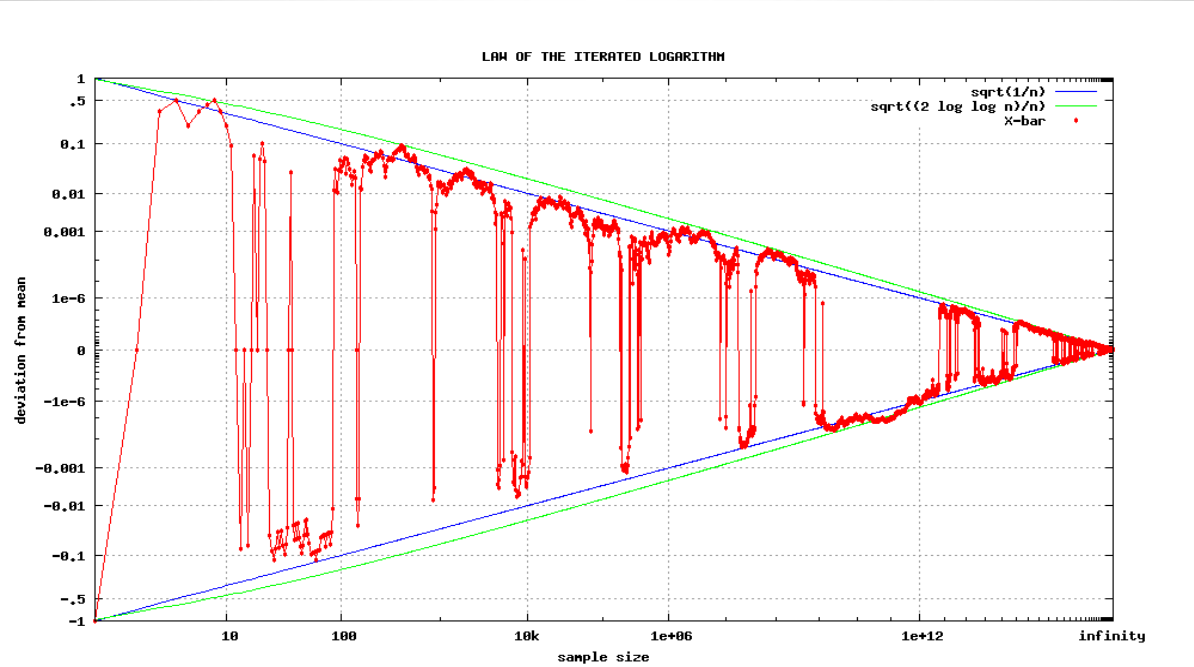
3. Закон повторного логарифма в общем виде

В более общем виде закон повторного логарифма можно сформулировать следующим образом. Если задана схема независимых испытаний Бернулли (p – вероятность положительного исхода в одном испытании, 1-p – вероятность отрицательного исхода), также справедлива формула:

$$\limsup_{n\to\infty}\frac{\frac{S_n-np}{\sqrt{np(1-p)}}}{\sqrt{2\ln\ln n}}=\limsup_{n\to\infty}\frac{S_n-np}{\sqrt{2\,np(1-p)\ln\ln n}}=1$$

4. Экспериментальная иллюстрация закона повторного логарифма

ЗПЛ занимает промежуточное положение между законом больших чисел и ЦПТ. Дальнейшие существенные продвижения в исследовании условий ЗПЛ связаны с работами А. Н. Колмогорова (1929) и В. Феллера (1943). Экспериментальная иллюстрация закона повторного логарифма представлена на рисунке



5. Статистические процедуры множественной проверки гипотез

Для проверки набора гипотез и увеличения точности обнаружения альтернативных гипотез при статистическом тестировании следует использовать определенный набор статистических критериев, а для принятия итогового решения – процедуры множественной проверки гипотез.

Одноэтапные процедуры, в первую очередь, предназначены для проверки качества тестируемой выборки. К недостаткам одноэтапных процедур можно отнести слабо поддающееся контролю уменьшение вероятности ошибки 1 рода и мощности. Двухэтапные процедуры позволяют увеличить мощность критерия и удерживать вероятность «ложной тревоги» на должном уровне.

6. Статистика закона повторного логарифма

Бинарную последовательность можно представить в виде реализации схемы Бернулли. Для $x \in \Sigma^n$ определим:

$$S(n) = \sum_{i=0}^{n-1} x_i ; \qquad S(n)^* = \frac{2S(n) - n}{\sqrt{n}} ,$$

где Σ^n – множество двоичных последовательностей длины n

Закон повторного логарифма дает оптимальную верхнюю оценку $\sqrt{2 \ln \ln n}$ для колебаний $S(n)^*$. Основываясь на этом факте, мы будем использовать следующую статистику:

$$S_{\text{зпл}}(n) = \frac{S(n)^*}{\sqrt{2 \ln \ln n}} = \frac{2S(n) - n}{\sqrt{2 n \ln n \ln n}} .$$

7. Распределение статистики закона повторного логприфма

Распределение, порожденное $S_{\text{зпл}}(n)$, определяет вероятностную меру на вещественной прямой R. Пусть $\mathcal{R} \in \Sigma^n$ – набор из m последовательностей со стандартным определением вероятности на нем. То есть, для каждой последовательности $x_0 \in \mathcal{R}$ положим $P[x = x_0] = {}^1/_m$. Тогда каждый набор $\mathcal{R} \in \Sigma^n$ порождает вероятностную меру $\mu_n^{\mathcal{R}}$ на R

$$\mu_n^{\mathcal{R}}(I) = P[S_{\text{зпл}}(n) \in I, x \in \mathcal{R}]$$

для каждого измеримого по Лебегу множества I на R. Для $\mathrm{U} = \Sigma^n$ введем μ_n^U – соответствующую вероятностную меру, порожденную равномерным распределением.

$$\mu_n^U\{(-\infty, z]\} = \Phi\left(z\sqrt{2\ln\ln n}\right) = \sqrt{2\ln\ln n}\int_{-\infty}^{z}\phi\left(y\sqrt{2\ln\ln n}\right)dy.$$

8. Процедура принятия решения

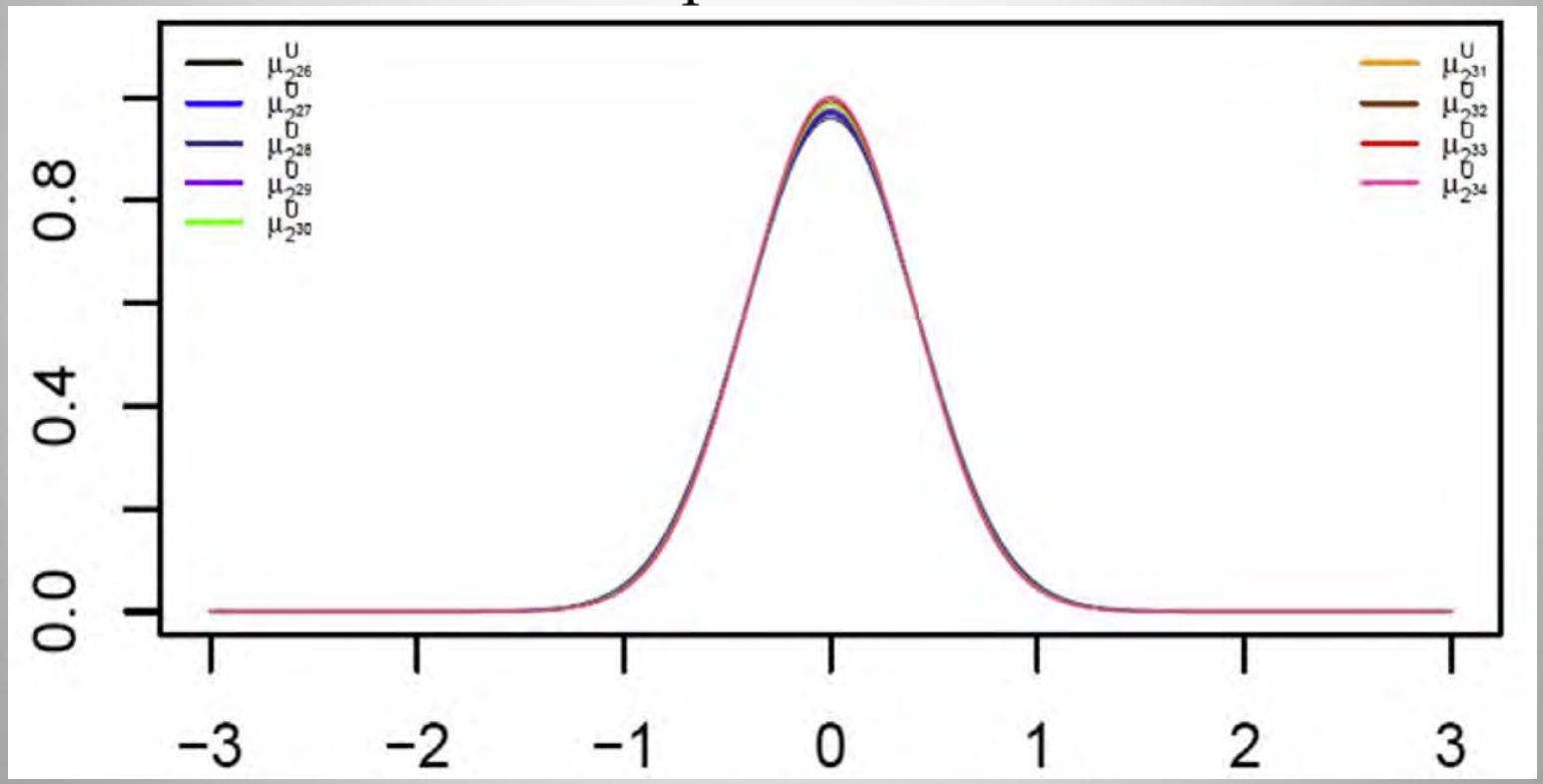
На первом этапе двухэтапной процедуры по каждой выборке вычисляется статистика критерия. На втором этапе по полученной последовательности значений статистик проверяется гипотеза согласия с теоретическим распределением статистики. Будем проверять гипотезу согласия последовательности значений статистики $S_{\text{зпл}}(n)$ с нормальным распределением на вещественной прямой R $(-\infty, \infty)$. Процедура принятия решения основана на критерии χ^2 согласия. Будем использовать в качестве дискретного разбиения вещественной прямой R множество $\mathcal{B}$, определяемое следующим образом:

$$\mathcal{B} = \cup I = \{(-\infty, -1), [1, \infty)\} \cup \{[0.05r - 1, 0.05r - 0.95)\}$$

$$(0 \le r \le 39)$$

8. График функции плотности теоретического распределения для различных объемов выборок

На основании данной формулы на рисунке приведены функции плотности распределений μ_n^U для различных объемов выборок



Функции плотности распределений μ_n^U

9. Этапы статистического тестирования с применением закона повторного логарифма

Чтобы оценить генератор G с применением закона повторного логарифма для теста Монобит, необходимо:

1. Осуществить генерацию набора $\mathcal{R} \in \Sigma^n$ из $m = 10\,000$ последовательностей возможно большей длины.
2. На первом этапе вычислить значения статистики $S_{\text{зпл}}(n)$ по всем m последовательностям.
3. На втором этапе сравнить между собой вероятностные меры $\mu_n^{\mathcal{R}_n}$ и μ_n^U по статистике χ^2 согласия [15]:

$$\chi^2 = \sum_{j=1}^{|\mathrm{B}|} \frac{\left[\nu_n^{\mathcal{R}_n}(I_j) - mp_n^{\mathrm{U}}(I_j)\right]^2}{mp_n^{\mathrm{U}}(I_j)},$$

где $\nu_n^{\mathcal{R}_n}(I_j)$ – частоты попадания значений статистики $S_{\text{зпл}}(n)$ в

9. Этапы статистического тестирования с применением закона повторного логарифма

Полагаем, что генератор G прошел тестирование по тесту Монобит с применением закона повторного логарифма, если P-значение статистики χ^2 согласия превышает заданный уровень значимости α, то есть, $P \geq \alpha$.

Для проверки гипотезы проведено тестирование 10000 последовательностей, выработанных соответственно линейным конгруэнтным генератором и физическим генератором «Ключ-ВС». Результаты сравнений выборок, полученных в результате применения закона повторного логарифма, с нормальным распределением по критерию χ^2 согласия приведены в таблице.

10. Тестирование последовательностей, выработанных ЛКГ и ФГ«Ключ-ВС»

Объем (GB)	ЛКГ		ФЗ «Ключ-ВС»	
	29 степ. свободы		29 степ. свободы	
	χ^2	P-знач.	$.\chi^2$	P-знач.
0.5 GB	29,97	0,4151	25,72	0,6402
1 GB	23,28	0,7631	30,88	0,3708
2 GB	18,82	0,9256	41,30	0,0647
3 GB	23.59	0,7488	30,22	0,4028
4 GB	30.37	0,3956	32,76	0,2876
5 GB	59.79	0,0006	22.78	0,7864
10 GB	98.98	$3{,}8\times10^{-10}$	27,40	0,4423

10. Тестирование последовательностей, выработанных ЛКГ и ФГ«Ключ-ВС»

Из таблицы видно, что для **всех** последовательностей, выработанных ФГ, выполняется гипотеза H_0 согласия с моделью независимых симметричных испытаний Бернулли на уровне значимости $\alpha = 0.01$. В то время как для последовательностей, выработанных ЛКГ, гипотеза H_0 верна только при объемах от 0.5 GB до 4 GB. Для последовательностей большего объема (5 GB, 10 GB) гипотеза H_0 не выполняется.

Причем с увеличением объема статистика χ^2 увеличивается. Хотя при этом P-значения статистик стандартного теста Монобит для всех объемов последовательностей согласуются с гипотезой H_0. Это означает, что уязвимости ЛКГ проявляются только в последовательностях, объемом не менее 5 GB и традиционным тестом Монобит не выявляются).

12. Тестирование последовательностей, выработанных в соответствии с СТБ 34.101.47-2012 (в режиме счетчика)

Проведено тестирование псевдослучайных последовательностей (объемом 5 GB и 10 GB), сгенерированных в соответствии со стандартом СТБ 34.101.47-2012 (в режиме счетчика).

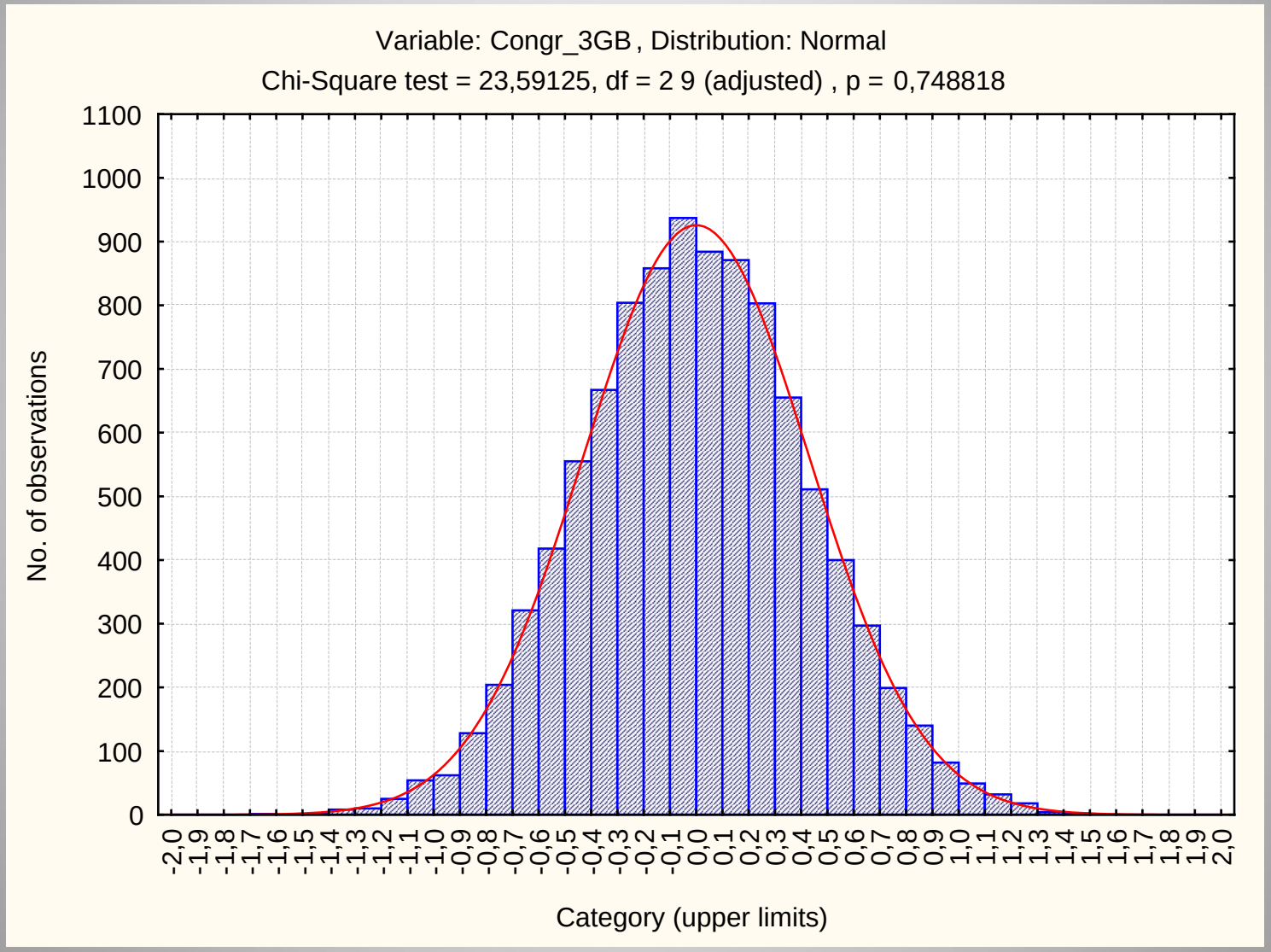
Объем (GB)	Степ. своб. (объед. групп)	χ^2	P-знач.	Степ. своб.	χ^2	P-знач.
5 GB	29	25,90	0,6303	41	37,03	0,6474
10 GB	27	23,36	0,6650	41	38,52	0,5812

Из таблицы видно, что для последовательностей (объемом 5 GB и 10 GB), сгенерированных в соответствии с СТБ 34.101.47-2012 (в режиме счетчика) выполняется гипотеза H_0 согласия с моделью независимых симметричных испытаний Бернулли на уровне значимости $\alpha = 0.01$. То есть данный алгоритм, в отличие от ЛКГ, является криптографически стойким при атаке с применением статистического расстояния и закона повторного логарифма для теста Монобит.

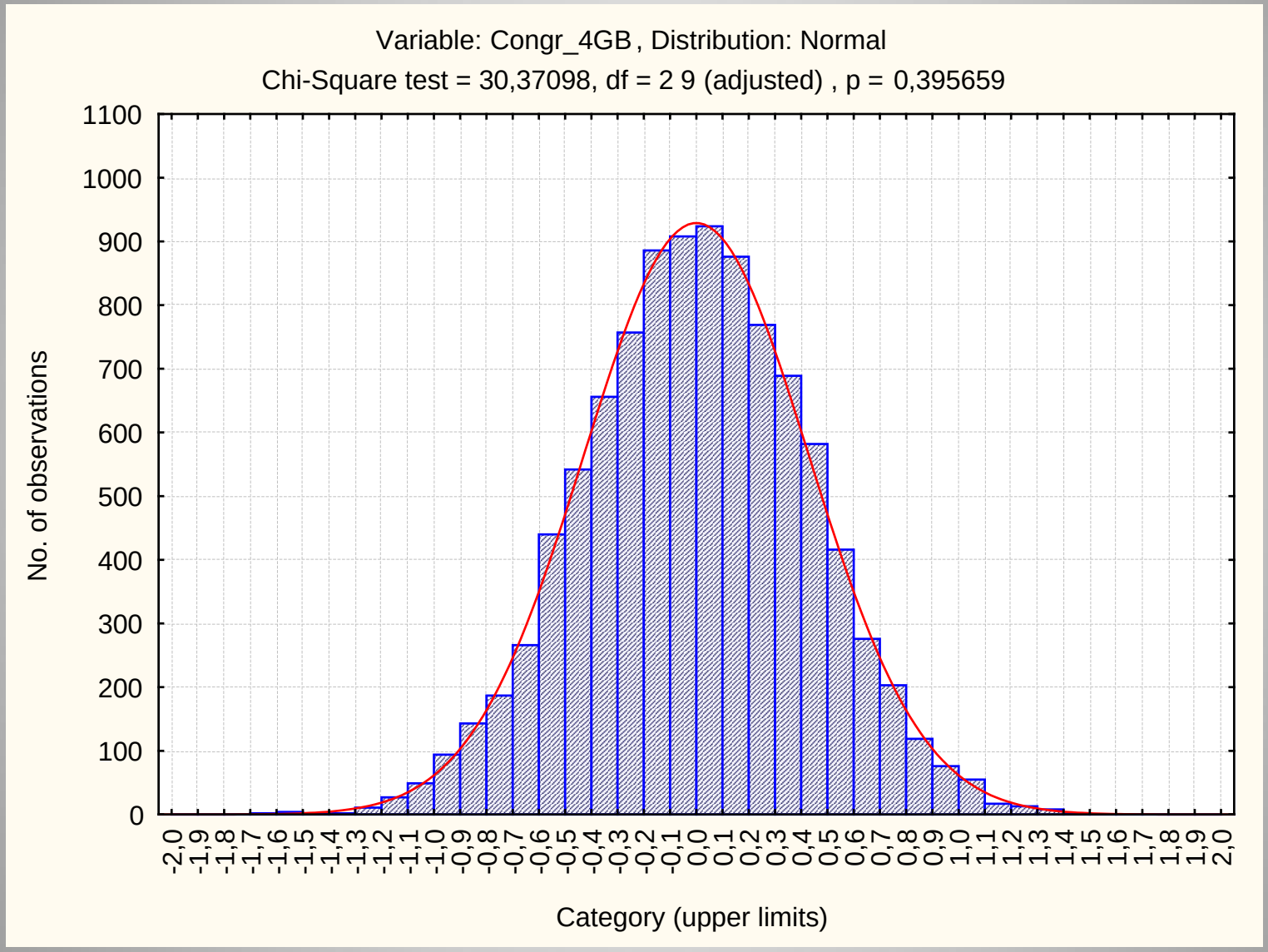
13. Результаты тестирования

Гистограммы частот выборок, полученных с применением закона повторного логарифма, для линейного конгруэнтного генератора, физического генератора «Ключ-ВС» и алгоритма генерации псевдослучайных последовательностей в соответствии с СТБ 34.101.47-2012, построенные с использованием программы «Статистика», приведены на следующих рисунках

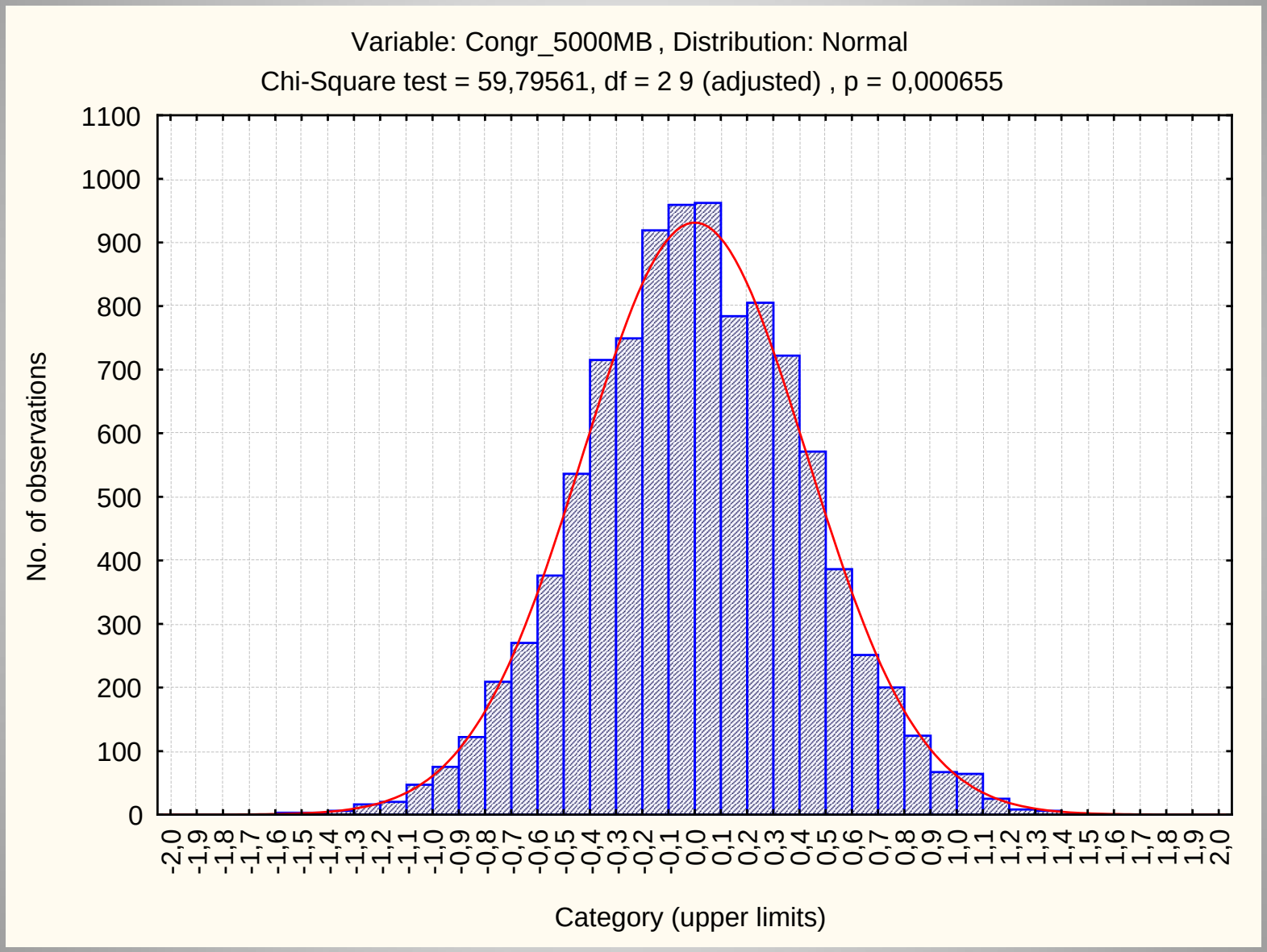
Гистограмма частот ЛКГ, 3 GB



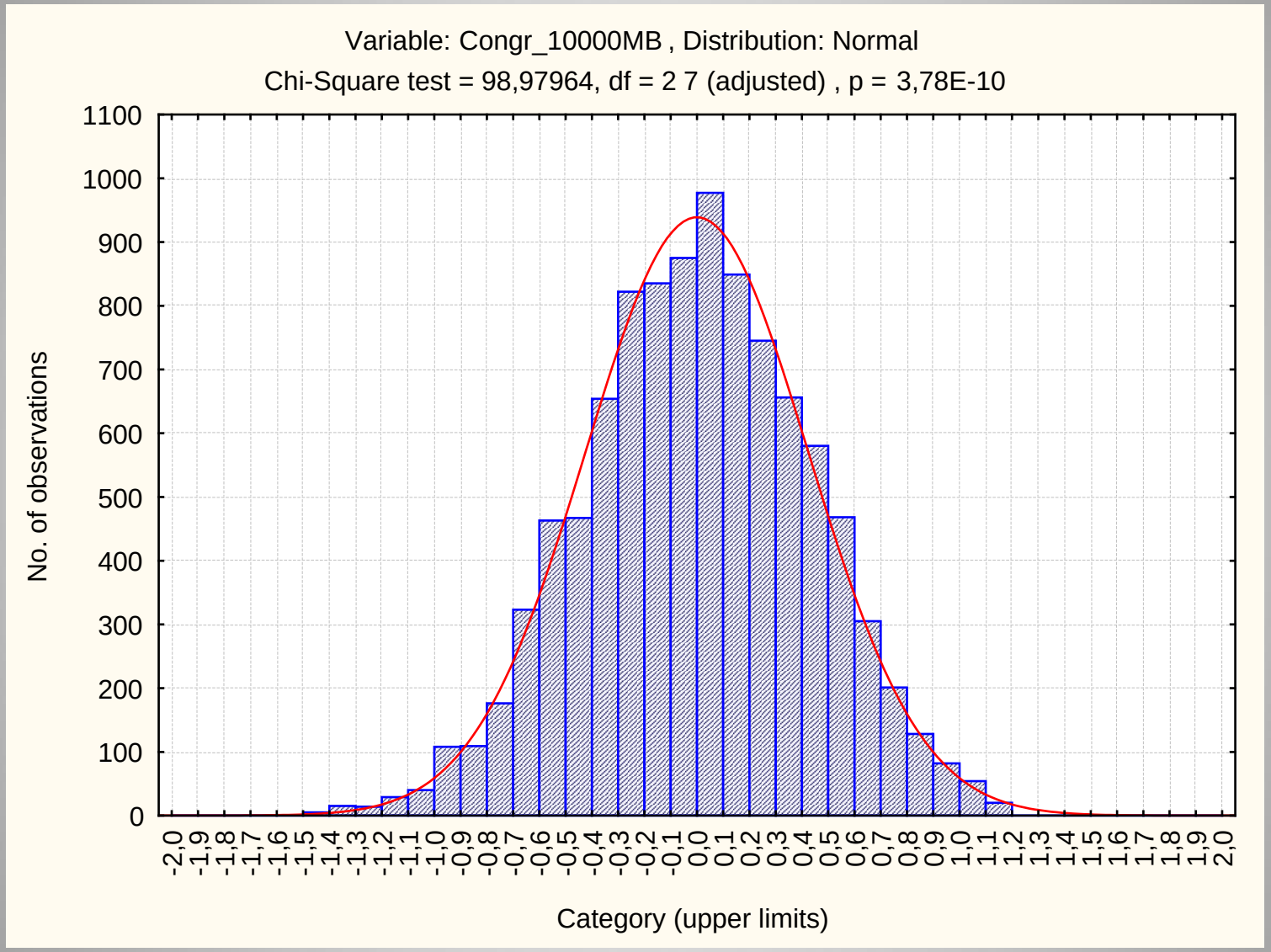
Гистограмма частот ЛКГ, 4 GB.



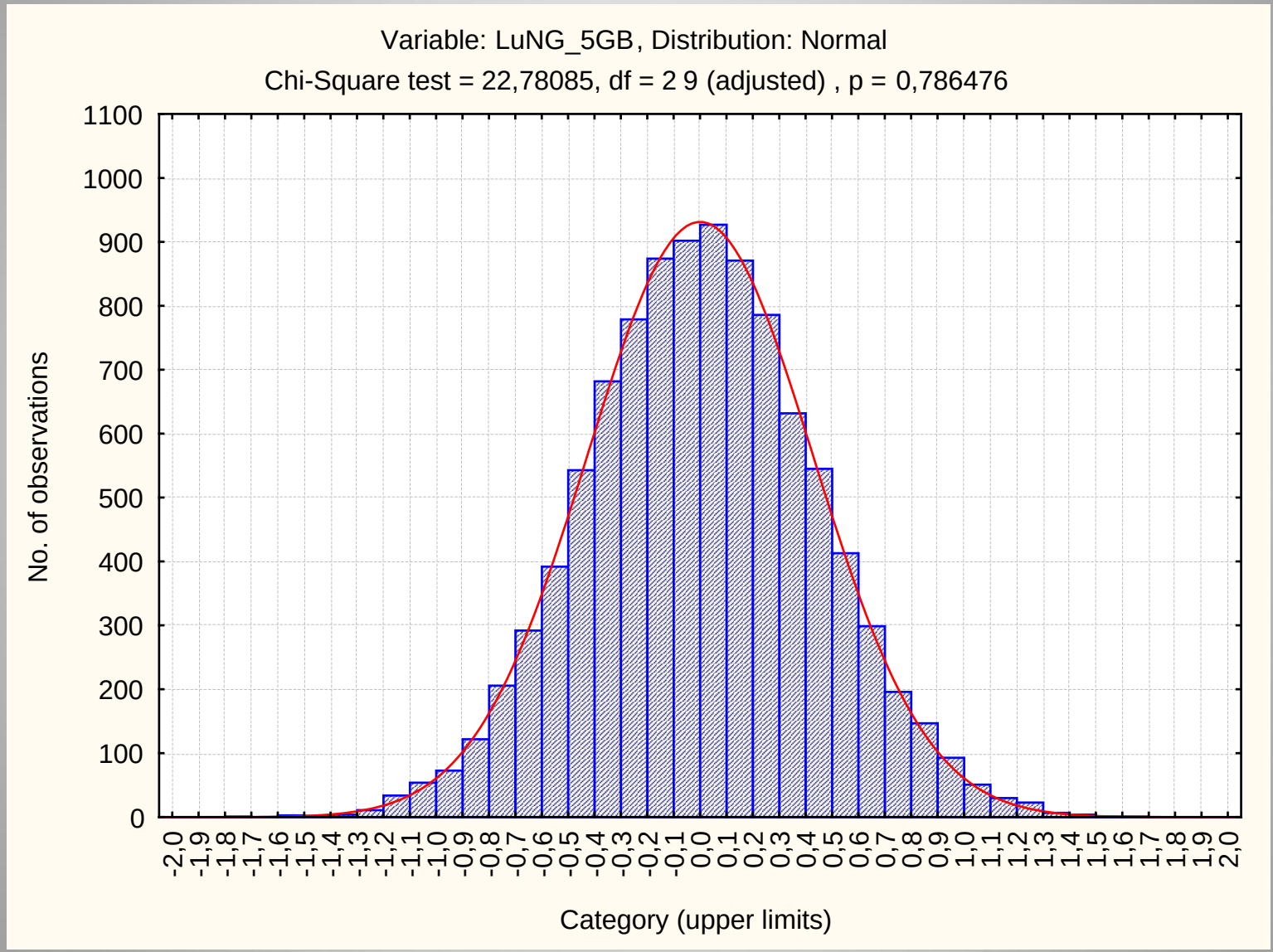
Гистограмма частот ЛКГ, 5 GB.



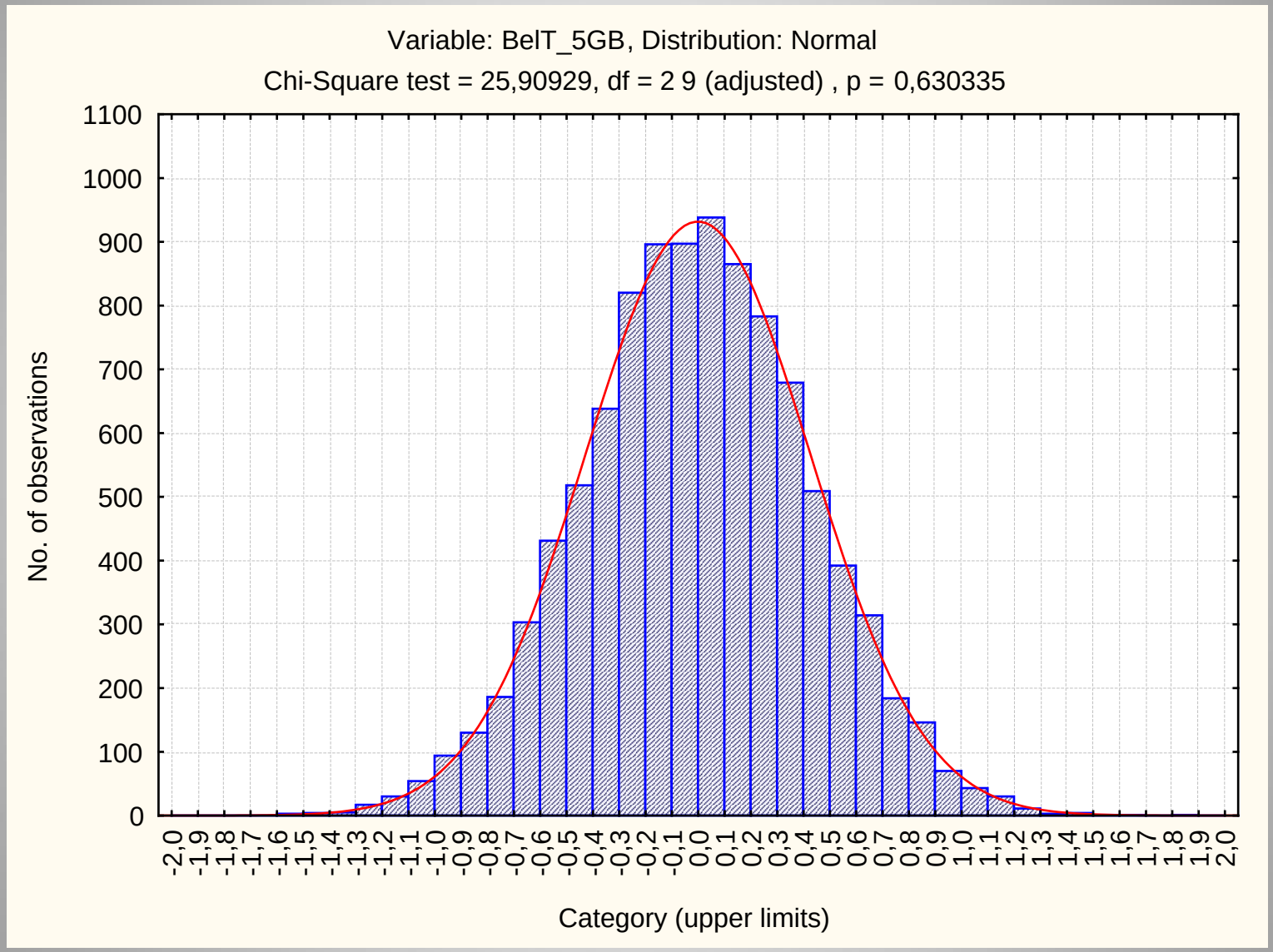
Гистограмма частот ЛКГ, 10 GB.



Гистограмма частот ФГ «Ключ-ВС», 5 GB



Гистограмма частот алгоритма СТБ 34.101.47-2012, 5 GB



ЗАКЛЮЧЕНИЕ

1. Установлено, что уязвимости в некоторых часто используемых реализациях генераторов псевдослучайных последовательностей не выявляются инструментами тестирования NIST SP800-22, в основном, из-за недостаточной длины тестируемых последовательностей.

2. Разработана двухэтапная процедура проверки гипотез с применением закона повторного логарифма для теста Монобит с использованием статистика хи-квадрат согласия.

3. Проведено тестирование последовательностей большого объема, сгенерированных линейным конгруэнтным генератором, а также физическим генератором на основе шумового диода «Ключ-ВС» по тесту Монобит. При этом физический генератор успешно прошел тестирование, а линейный конгруэнтный генератор тестирование провалил.

4. Тестирование с применением статистического расстояния и закона повторного логарифма показало, что алгоритм генерации псевдослучайных последовательностей согласно СТБ 34.101.47-2012 (в режиме счетчика) является криптографически стойким при атаке с применением закона повторного логарифма для теста Монобит.

Спасибо за внимание!