XXIII НАУЧНО-ПРАКТИЧЕСКАЯ КОНФЕРЕНЦИЯ
«КОМПЛЕКСНАЯ ЗАЩИТА ИНФОРМАЦИИ»,
г. СУЗДАЛЬ, 22 – 24 мая 2018 года

ПОСТРОЕНИЕ ТЕОРЕТИКО-ВЕРОЯТНОСТНОЙ МОДЕЛИ ФИЗИЧЕСКОГО ПРОЦЕССА НА ОСНОВЕ ШУМОВОГО ДИОДА

НИИ прикладных проблем математики и информатики БГУ

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

ВВЕДЕНИЕ

Согласно **Принципам разработки и модернизации шифровальных (криптографических) средств защиты информации (ТК 26)** для физических датчиков случайных числовых последовательностей (ФДСЧП), входящих в состав средств криптографической защиты информации, должна быть разработана теоретико-вероятностная модель (ТВМ) используемого в датчике случайного физического процесса и проведена экспериментальная проверка соответствия указанной модели реализации соответствующего ФДСЧП [1].

Таким образом, статистическое моделирование – это численный метод решения математических задач, при котором искомые величины представляются вероятностными характеристиками какого-либо случайного процесса. Данный процесс моделируется, после чего нужные характеристики приближённо определяются путём статистической обработки «наблюдений» модели.

ТВМ – это модель, основанная на применении статистической или теоретико-вероятностной методологии по отношению к повторяющимся феноменам, в которой обеспечивается учет случайных факторов в процессе функционирования системы.

ВВЕДЕНИЕ

Данная модель оперирует количественными критериями при оценке повторяющихся явлений и позволяет учитывать их нелинейность, динамику, случайные возмущения за счет выдвижения на основе анализа результатов наблюдений гипотез о характере распределения некоторых случайных величин, сказывающихся на поведении системы.

По существу, теоретико-вероятностные и статистические модели отличаются уровнем неопределенности знаний о моделируемой системе, существующей на момент синтеза модели. В случае, когда представления о системе основываются исключительно на гипотезах о характере системы и возмущающих воздействий, не подкрепленных результатами наблюдений, ТВМ является единственно возможной. Если на этапе синтеза модели уже существуют данные, полученные опытным путем, то появляется возможность подкрепления гипотез за счет их статистической обработки.

1. Гипотетические вероятностные функции распределения физического источника случайности на основе шумового диода

Наиболее часто в ДСЧП в качестве первичного источника энтропии используется физический источник случайности на основе шумового диода. Работа диода основана на туннельном пробое обратно смещенного **(*p*-*n*)** – перехода полупроводникового диода. Аналоговый сигнал, через конденсатор подается на вход компаратора. Выход компаратора подается на вход таймера, работающего в режиме счетчика. Таймер производит подсчет количества импульсов с выхода компаратора на заданном временном интервале. В качестве случайного бита принимается младший бит (0 или 1) счетчика таймера.

Результаты повторных измерений случайной величины ξ (числа отсчетов) могут значительно различаться. Количество отсчетов с выхода компаратора во временном интервале является случайным событием. Необходимо определить вероятность $P_k(t)$, что в интервале времени t счетчик зарегистрирует k импульсов.

1. Гипотетические вероятностные функции распределения физического источника случайности на основе шумового диода

Если для ξ, выполняются условия:

а) случайная величина ξ может принимать только целые положительные значения, включая 0;

б) вероятность двух (и более) событий на достаточно малом временном или пространственном интервале бесконечно мала по сравнению с вероятностью одного события;

в) события статистически независимы (во времени / пространстве);

г) время (или пространство) однородно для изучаемых событий,

то ξ имеет распределение Пуассона (распределение дискретного типа).

Можно выдвинуть гипотезу, что случайная величина ξ - число импульсов k с выхода компаратора шумового диода на заданном временном интервале t, также должна иметь распределение Пуассона:

$$P_k(t) = \frac{(nt)^k}{k!} e^{-nt} = \frac{\lambda^k}{k!} e^{-\lambda}, \qquad (1)$$

где $\lambda = nt$, n – среднее число импульсов за единицу времени, то есть, их интенсивность.

1. Гипотетические вероятностные функции распределения физического источника случайности на основе шумового диода

Математическое ожидание (среднее количество отсчетов) и дисперсия равны: $M(\xi) = nt = \lambda;\quad D(\xi) = M(\xi) = nt = \lambda$. Т.е, распределение Пуассона определяется заданием одного параметра – среднего числа отсчетов.

Условия формирования распределения Пуассона могут нарушаться. Например, парное прохождение импульсов нарушает условия *а*) – *г*).

Кроме того, любое устройство затрачивает на измерение и регистрацию события время, в течение которого оно не способно «правильно» обработать следующее событие. Это так называемое мертвое время (dead time). Влияние мертвого времени нарушает условия *в*), *г*). В этом случае ξ будет иметь другое распределение.

С ростом ***k*** распределение становится симметричным. При $\sqrt{\lambda} \gg 1$ – становится практически полностью симметричным. В этом случае целесообразно рассматривать вероятность попадания ***k*** в заданный интервал Δk вблизи некоторого значения ***k***. Т. е. совершается переход от дискретного распределения к непрерывному. Распределение Пуассона переходит в нормальное распределение, для которого дисперсия равна математическому ожиданию.

2. Проверка гипотез о законе распределения с применением критерия согласия хи-квадрат

Предположим, что некоторый алгоритм осуществляет моделирование случайной величины ξ (например, количество отсчетов с выхода компаратора на заданном временном интервале). В результате *n*-кратного обращения к данному алгоритму моделируется случайная выборка $X = \{x_1, ..., x_n\}$. Необходимо проверить гипотезу о том, что случайная величина ξ имеет функцию распределения $F_\xi(x) = F_0(x)$, где $F_0(x)$ – фиксированная функция распределения при некоторых неизвестных значениях параметров $a_j \ (j = 1, ..., s)$

Предположим, что выборка разбита на *r* групп, соответствующих *r* непересекающимся множествам $S_1, ..., S_r$. Обозначим наблюдаемую группу частот $\nu_1, ..., \nu_r$, а соответствующие вероятности $p_i(a_1, ..., a_s)$

Если бы значения параметров были известны, то можно было бы применить критерий хи-квадрат:

$$\chi^2 = \sum_{i=1}^{r} \frac{\left[\nu_i - np_i(a_1, ..., a_s)\right]^2}{np_i(a_1, ..., a_s)} \qquad (2)$$

2. Проверка гипотез о законе распределения с применением критерия согласия хи-квадрат

Однако в данном случае значения параметров a_j неизвестны и должны быть оценены по выборке. Очевидно, что свойства выборочного распределения статистики χ^2 будут в той или иной степени зависеть от избранного метода. В частности при применении метода оценки по минимуму необходимо определить «наилучшие» значения параметров a_j так, чтобы сделать величину χ^2 сколь угодно малой. Доказано, что χ^2 при подстановке a_j в формулу (2), при $n \to \infty$ имеет распределение хи-квадрат с $r–s–1$ степенями свободы [2].

Исследования проводились на суммарной выборке объемом 10 240 000 отсчетов. Суммарная выборка была разбита на 10 выборок объемом 1 024 000 отсчетов. Проверка гипотез о функции распределения осуществлялась для выборок объемом соответственно: 10 240 000 отсчетов и 3 072 000 отсчетов (последние 3 выборки объемом 1 024000 отсчетов каждая). В таблице 1 приведены оценки параметров – математического ожидания и дисперсии.

2. Проверка гипотез о законе распределения с применением критерия согласия хи-квадрат

Таблица 1 – Оценки параметров для выборок различного объема

Объем выборки	10 240 000 отсчетов	3 072 000 отсчетов
$\hat{m} = \frac{1}{n}\sum_i \nu_i \xi_i$	43,63	44,15
$\hat{\sigma}^2 = \frac{1}{n-1}\sum_i \nu_i (\xi_i - \hat{m})^2$	14,25	14,29

Рассмотрим применение критерия хи-квадрат для различных гипотез о предполагаемом распределении случайной величины ξ, представляющей собой количество импульсов с выхода компаратора на заданном временном интервале. В качестве возможных функций распределения рассмотрим **распределение Пуассона**, **отрицательное биномиальное распределение**, **нормальное распределение**.

2.1. Распределение Пуассона. В случае справедливости гипотезы математическое ожидание и дисперсия должны быть равны. Однако из таблицы видно, что оценки математического ожидания и дисперсии значительно различаются. Следовательно, гипотеза отвергается.

2. Проверка гипотез о законе распределения с применением критерия согласия хи-квадрат

2.2. Отрицательное биномиальное распределение. Если выборки обнаруживают значимое отклонение от распределения Пуассона, то совпадение можно значительно улучшить, выдвинув гипотезу, что параметр является случайной величиной с плотностью вероятности

$$a^x / \Gamma(x)\, x^{x-1} e^{-ax}$$

где ***a, x*** – положительные параметры. В общем случае выборки описываются отрицательным биномиальным распределением. В случае справедливости гипотезы между оценками математического ожидания и дисперсии должно соблюдаться соотношение:

$$\hat{\sigma}^2 = \hat{m}/p \text{, где } 0 \le p \le 1$$

То есть дисперсия должна быть не меньше математического ожидания. Однако из таблицы 1 видно, что условия не выполняются. Таким образом, гипотеза отклоняется.

2. Проверка гипотез о законе распределения с применением критерия согласия хи-квадрат

2.3. Нормальное распределение. Сравниваем теоретические и опытные частоты с помощью критерия хи-квадрат согласия по формуле (2), задаем уровень значимости α и определяем число степеней свободы $k = r–3$, где r – число групп после объединения. С использованием программного комплекса STATISTICA, на основе полученных выборок были построены гистограммы. На рисунке 1 приведена гистограмма частот для выборки 3 072 000 отсчетов.

Из гистограммы видно, что она имеет одну ярко выраженную вершину. Распределение частот достаточно симметрично. При этом не удалось подтвердить гипотезу о нормальном распределении на приемлемом уровне значимости, хотя значение критерия хи-квадрат становится существенно меньше, чем для выборки объемом 10 240 000 отсчетов. Это означает, что данная выборка в большей степени согласуется с гипотезой о нормальном распределении, чем суммарная выборка. Возможно, процесс функционирования датчика с течением времени стабилизировался (дальнейшее нагревание диода существенно замедлилось).

2. Проверка гипотез о законе распределения с применением критерия согласия хи-квадрат

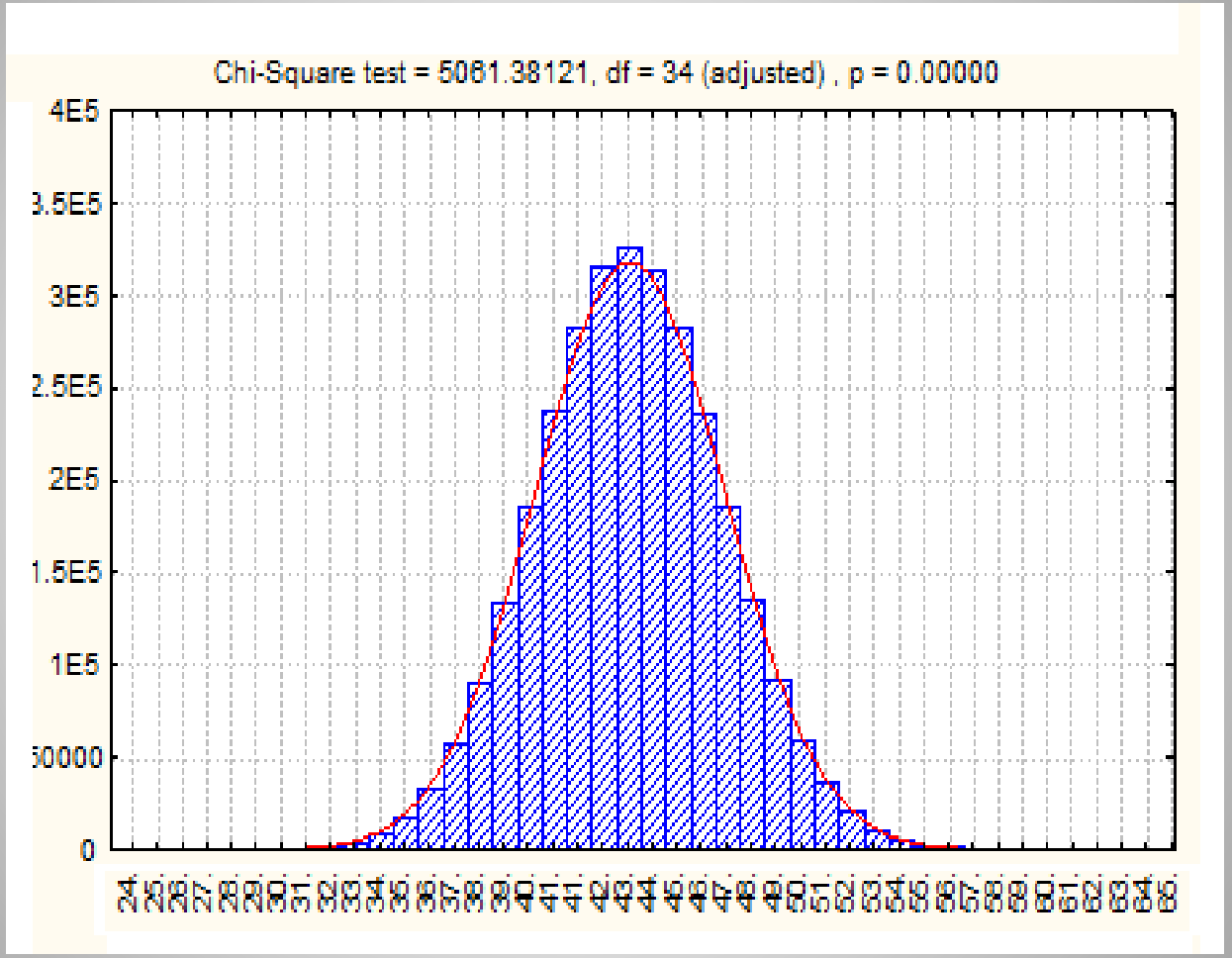

Рисунок 1 – Гистограмма частот выборки объемом 3 072 000 отсчетов

3. Оценка стабильности вероятностных характеристик физического процесса, используемого в ДСЧП

Рассмотрим гипотезу о том, что распределение вероятностей случайной величины зависит от времени, и определим статистическую значимость отклонений его параметров с течением времени.

Предположим, что имеем комплект из $M \geq 2$ первичных последовательностей, выработанных датчиком в промежутки времени $\Delta t_1, ..., \Delta t_M$. Из комплекта сформируем M независимых выборок. Каждая из выборок есть реализация полиномиальной схемы с N исходами. Объемы выборок равны $n_1, ..., n_M$. Обозначим $n = \sum_{k=1}^{M} n_k$ $n_0 = min\{n_1, ..., n_M\}$, $\nu_{k,i}$ – частота символа i.

Гипотеза однородности предполагает, что вероятностные свойства наблюдаемой последовательности не изменяются во времени. Т. е, во все промежутки времени $\Delta t_1, ..., \Delta t_M$ вероятности $p_{1,i} = ... = p_{M,i}$ исходов **1,…,*N***, распределенные по определенному закону (например, по закону Пуассона, Гаусса или иному), остаются неизменными.

3. Оценка стабильности вероятностных характеристик физического процесса, используемого в ДСЧП

Полагаем, что выборки статистически однородны и для них выполняется гипотеза H_0 , если $p_{1,i} = ... = p_{M,i}$ В противном случае будем говорить, что выполняется альтернатива H_1. Для проверки однородности выборок будем использовать модификацию статистики хи-квадрат, предложенную А.М. Зубковым [3]:

$$\zeta^2 = \inf_{p_1,...,p_N>0} n\sum_{k=1}^{M}\sum_{i=1}^{N}\frac{1}{n_k m_i}\left(\nu_{k,i} - p_i\right)^2 = \left(\sum_{i=1}^{N}\sqrt{\sum_{k=1}^{M}\frac{\nu_{k,i}^2}{n_k}}\right)^2 \qquad (3)$$

где $m_i = \sum_{k=1}^{M}\nu_{k,i}$

Если справедлива гипотеза H_0, то при $n_0 \to \infty$ статистика $\zeta^2 - n$ сходится к распределению хи-квадрат с $(N-1)(M-1)$ степенями свободы.

3. Оценка стабильности вероятностных характеристик физического процесса, используемого в ДСЧП

Для анализа был взят комплект из 10 выборок по 1 024 000 отсчетов каждая. По формуле (3) осуществлялась оценка однородности комплекта в целом. В этом случае $\zeta^2 - n$ = **153 953.321** для $9 \times 30 = 270$ степеней свободы. Это означает, что выборка в целом чрезвычайно неоднородна.

Проводилось также попарное сравнение выборок на однородность по формуле (3). Число степеней свободы: $1 \times 30 = 30$.

При попарных сравнениях также отмечается неоднородность выборок. Чем больше разность между номерами выборок (i, j), тем больше статистическое различие между ними. Выборки, зарегистрированные в соседние промежутки времени, более однородны, чем выборки, выработанные через более длительные временные отрезки.

4. Оценка наличия марковской зависимости

Однородная цепь Маркова ***s***-го порядка **(*s* < ∞)** описывает зависимость каждого наблюдения только от *s* предыдущих состояний. С ростом *s* число параметров цепи Маркова порядка *s* растет с экспоненциальной скоростью (порядка N^{s+1}), что ограничивает применение этой модели небольшими значениями *s*.

Для выявления марковской зависимости в анализируемой последовательности строилась статистическая оценка порядка цепи Маркова $\hat{s}$. Значение $\hat{s} > 0$ свидетельствует о наличии марковской зависимости, значение $\hat{s} = 0$ соответствует последовательности независимых испытаний.

Для построения оценок порядка *s* использовался информационный функционал Байеса (BIC), который учитывает число параметров модели [4]. Он имеет вид: BIC(s) = $-2\hat{l}(X,s) + D\log(n)$, где $\hat{l}(X,s)$ – статистическая оценка логарифмической функции правдоподобия, вычисленная по последовательности $X = (x_1,..., x_n)$ в предположении, что порядок цепи Маркова равен *s*, $D = N^s(N–1)$ – число независимых параметров модели.

4. Оценка наличия марковской зависимости

Оценка $\hat{s}$ определяется решением задачи на минимум:

$$\hat{s} = \arg \min_{0 \le s' \le S} BIC(s')$$

где S – максимально допустимое значение порядка s, задаваемое априорно исходя из имеющихся данных. По анализируемой последовательности X были вычислены значения BIC для s =0, 1, 2, 3. В результате оценивания порядка для модели однородной цепи Маркова на основе байесовского информационного критерия было получено $\hat{s} = 0$, которому соответствует минимум ***BIC***. Это означает, что по критерию принято решение о несогласии последовательности с моделью однородной бинарной цепи Маркова, т.е. марковская зависимость в последовательности не обнаружена.

5. Оценка статистических свойств выходной бинарной последовательности

Выходная бинарная последовательность получена из первичной последовательности посредством определения значения младшего бита счетчика числа отсчетов (0 – четное число, 1 – нечетное число). Вычислим гипотетические вероятности 0 и 1 в выходной бинарной последовательности. В таблице 2 приведено количество четных и нечетных отсчетов в гистограммах частот выборок, а также оценки их согласия с равновероятным распределением по критерию хи-квадрат.

Таблица 2 – Распределение четных и нечетных отсчетов в выборках

Объем выборки	**10 240 000 отсчетов**	**3 072 000 отсчетов**
Число нечетных отсчетов	5 121 613	1 536 263
Число нечетных отсчетов	5 118 387	1 535 637
P-значения статистики χ^2 (с одной степенью свободы)	0.313395	0.720967

Очевидно, что принимается гипотеза о равновероятном распределении четных и нечетных отсчетов. Следовательно, в бинарных последовательностях распределение 0 и 1 будет близко к равновероятному.

ЗАКЛЮЧЕНИЕ

1. Физический процесс, генерируемый источником случайности на основе шумового диода, является недостаточно стационарным.

2. Частоты встречаемости отсчетов достаточно симметрично распределены относительно математического ожидания и имеют одну моду. При этом не удалось подтвердить гипотезу о нормальном распределении на приемлемом уровне значимости.

3. Наличие марковской зависимости с применением байесовского информационного критерия (BIC) не выявлено.

4. Вероятности отсчетов согласуются с гипотезой о равновероятном распределении четных и нечетных отсчетов. Следовательно, в выходных бинарных последовательностях распределение 0 и 1 также будет близко к равновероятному.

Спасибо за внимание!