**ПРИМЕР ПРЕДСТАВЛЕНИЯ ТЕЗИСОВ**

*УДК 004.056*

**ПОСТРОЕНИЕ ТЕОРЕТИКО-ВЕРОЯТНОСТНОЙ МОДЕЛИ ФИЗИЧЕСКОГО ПРОЦЕССА НА ОСНОВЕ ШУМОВОГО ДИОДА, ИСПОЛЬЗУЕМОГО В ДАТЧИКЕ СЛУЧАЙНЫХ ЧИСЛОВЫХ ПОСЛЕДОВАТЕЛЬНОСТЕЙ**

А. И. ТРУБЕЙ, М.В. МАЛЬЦЕВ, И.К. ПИРШТУК, В.Ю. ПАЛУХА

*Учреждение БГУ «НИИ прикладных проблем математики и информатики»,
 г. Минск, 220030, Республика Беларусь*

**Введение**

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].

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. 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 случайных событий физического источника случайности на основе шумового диода**

Наиболее часто в ДСЧП в качестве первичного источника энтропии используется физический источник случайности на основе шумового диода. В частности, в физическом источнике случайности на основе шумового 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  (числа отсчетов) могут значительно различаться. Количество отсчетов с выхода компаратора во временном интервале является случайным событием. Необходимо определить вероятность , что в интервале времени  счетчик зарегистрирует  импульсов. Если для , представляющей собой число событий за фиксированное время, выполняются условия:

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

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

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

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

то  имеет распределение Пуассона (распределение дискретного типа).

Таким образом, можно выдвинуть гипотезу, что случайная величина , представляющая собой число импульсов с выхода компаратора шумового диода на заданном временном интервале , также должна иметь распределение Пуассона:

 (1)

где  – среднее число импульсов за единицу времени, то есть, их интенсивность.

Математическое ожидание (среднее количество отсчетов) и дисперсия равны:  . То есть, распределение Пуассона полностью определяется заданием только одного параметра – среднего количества отсчетов.

Условия формирования распределения Пуассона, указанные в п.п. а) – г), иногда могут нарушаться. Например, парное прохождение импульсов нарушает условия а) – г). Кроме того, любое устройство затрачивает на измерение и регистрацию события конечное 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  будет иметь другое распределение.

Анализ формулы (1) показывает, что с ростом  распределение становится симметричным. При  – становится практически полностью симметричным. Условие  означает, что вероятности близких значений  будут почти одинаковы, и в этом случае целесообразно рассматривать вероятность не отдельного значения , а вероятность попадания  в заданный интервал значений  вблизи некоторого значения . Тем самым совершается переход от дискретного распределения к непрерывному. При больших значениях  распределение Пуассона переходит в нормальное распределение, для которого дисперсия равна математическому ожиданию.

**2. Проверка гипотез о законе распределения с применением критерия согласия хи-квадрат, когда по выборке оцениваются параметры распределения**

Предположим, что некоторый алгоритм осуществляет моделирование случайной величины  (например, количество отсчетов с выхода компаратора на заданном временном интервале). В результате -кратного обращения к данному алгоритму моделируется случайная выборка . Необходимо при помощи  проверить гипотезу  о том, что случайная величина  имеет функцию распределения , где  – фиксированная функция распределения при некоторых неизвестных значениях параметров .

Предположим, что выборка разбита на  групп, соответствующих  непересекающимся множествам . Обозначим наблюдаемую группу частот , а соответствующие вероятности – , . Если бы значения параметров были известны, то можно было бы применить критерий хи-квадрат:

. (2)

Однако в данном случае значения параметров $α\_{j}$ неизвестны и должны быть оценены по выборке. Существует множество различных методов оценки параметров
, так что свойства выборочного распределения статистики  будут в той или иной степени зависеть от избранного метода. В частности при применении метода оценки по минимуму  необходимо определить «наилучшие» значения параметров  так, чтобы сделать величину  сколь угодно малой. Доказано, что величина  при подстановке этих значений $α\_{j}$ в формулу (2), при  имеет распределение хи-квадрат с  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. На основании эмпирического анализа частот встречаемости отсчетов можно сделать вывод, что значения частот, в том числе средние значения, с течением времени постепенно смещаются в сторону увеличения. Это означает, что физический процесс, генерируемый источником случайности на основе шумового диода, является недостаточно стационарным (возможно, за счет постепенного нагревания диода).

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.

Таблица 1 –  Оценки параметров для выборок различного объема

|  |  |  |
| --- | --- | --- |
| Объем выборки | 10 240 000 отсчетов | 3 072 000 отсчетов |
| Мат. ожидание:   | 43,63 | 44,15 |
| Дисперсия:   | 14,25 | 14,29 |

Рассмотрим применение критерия хи-квадрат для различных гипотез о предполагаемом распределении случайной величины , 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 значения математического ожидания и дисперсии должны быть равны. Однако из таблицы 1 видно, что оценки математического ожидания и дисперсии значительно различаются. Следовательно, гипотеза отвергается.

**2.2. Отрицательное биномиальное распределение.** Если выборки обнаруживают значимое отклонение от распределения Пуассона, то совпадение можно значительно улучшить, выдвинув гипотезу о том, что параметр  является случайной величиной с плотностью вероятности , где  – положительные параметры. В общем случае выборки описываются отрицательным биномиальным распределением. В случае справедливости гипотезы между оценками математического ожидания и дисперсии должно соблюдаться соотношение , где . То есть дисперсия должна быть не меньше математического ожидания. Однако из таблицы 1 видно, что условия не выполняются. Таким образом, гипотеза отклоняется.

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



Рисунок 1 – Гистограмма частот выборки объемом 3 072 000 отсчетов

Из гистограммы, приведенной на рисунке 1, видно, что она имеет одну ярко выраженную вершину. Распределение частот отсчетов достаточно симметрично. При этом также не удалось подтвердить гипотезу о нормальном распределении на приемлемом уровне значимости, хотя значение критерия  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).

**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  первичных последовательностей, выработанных датчиком в промежутки времени . Из данного комплекта сформируем  независимых выборок. Каждая из выборок есть реализация полиномиальной схемы с  исходами. Объемы выборок равны . Обозначим , ,  – частота символа .

Гипотеза однородности предполагает, что вероятностные свойства наблюдаемой последовательности не изменяются во времени. То есть, во все промежутки времени  вероятности  исходов , распределенные по определенному закону (например, по закону Пуассона, Гаусса или иному), остаются неизменными.

Полагаем, что выборки статистически однородны и для них выполняется гипотеза , если ; . В противном случае будем говорить, что выполняется альтернатива . Для проверки однородности выборок будем использовать модификацию статистики хи-квадрат, предложенную А.М. Зубковым [3]:

, где . (3)

Если справедлива гипотеза  то при  статистика  сходится к распределению хи-квадрат с степенями свободы.

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 =152 9 для 9×30=270 степеней свободы. Это означает, что выборка в целом чрезвычайно неоднородна. Проводилось также попарное сравнение выборок на однородность по формуле (3). Результаты сравнений приведены в таблице 2. На пересечении -й строки и -го столбца представлены значения статистики , вычисленные при сравнении -й и -й выборок. Число степеней свободы:1×30=30.

Таблица 2 – Проверка однородности выборок при попарном сравнении

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| № выборки | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
| 1 | 220 | 3943 | 9843 | 27209 | 32323 | 30654 | 49324 | 50778 | 77230 |
| 2 |  | 2992 | 7097 | 24926 | 29589 | 37445 | 35697 | 56683 | 72582 |
| 3 |  |  | 2776 | 6731 | 9852 | 25557 | 31870 | 30682 | 53979 |
| 4 |  |  |  | 2316 | 3938 | 7520 | 21979 | 28635 | 39314 |
| 5 |  |  |  |  | 484 | 3216 | 5057 | 10819 | 29190 |
| 6 |  |  |  |  |  | 869 | 3780 | 7402 | 24503 |
| 7 |  |  |  |  |  |  | 756 | 3939 | 9139 |
| 8 |  |  |  |  |  |  |  | 959 | 5379 |
| 9 |  |  |  |  |  |  |  |  | 2499 |

При попарных сравнениях также отмечается неоднородность выборок. Чем больше разность между номерами выборок , тем больше статистическое различие между ними. Выборки, зарегистрированные в соседние промежутки времени, более однородны, чем выборки, выработанные через более длительные временные отрезки. Таким образом, подтверждаются результаты эмпирического анализа выборок.

**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 *Ns*+1), что ограничивает применение этой модели небольшими значениями *s*.

Для выявления марковской зависимости в анализируемой последовательности строилась статистическая оценка порядка цепи Маркова . Значение говорит о наличии марковской зависимости, значение соответствует последовательности независимых испытаний.

Метод максимального правдоподобия, который традиционно используется для построения статистических оценок, не применим для оценивания порядка цепи Маркова. Использование данного метода для решения этой задачи приводит к проблеме, известной как over-fitting – чрезмерной «подгонке» модели под имеющиеся данные. В результате выбирается наиболее сложная модель, поскольку за счет увеличения числа параметров становится возможным увеличивать функцию правдоподобия. Поэтому для построения оценок порядка *s* использовался информационный функционал Байеса (BIC), который учитывает число параметров модели [4]. Он имеет вид: BIC(s) = , где  – статистическая оценка логарифмической функции правдоподобия, вычисленная по последовательности *X* = (*x*1,..., *xn*) в предположении, что порядок цепи Маркова равен *s*, *D* = *Ns*(*N*–1) – число независимых параметров модели.

Оценка  определяется при решении задачи на минимум: , где *S* – максимально допустимое значение порядка *s*, задаваемое априорно исходя из имеющихся данных. В таблице 3 приведены значения BIC для *s*= 0, 1, 2, 3, вычисленные по анализируемой последовательности *X*.

Таблица 3 – Значения BIC для последовательности *X* объемом 10 240 000 отсчетов

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| *s* | 0 | 1 | 2 | 3 |
| BIC(*s*) | 28093850.4 | 28103693.1 | 28580777.7 | 48050062.4 |

Таким образом, в результате оценивания порядка для модели однородной цепи Маркова на основе байесовского информационного критерия было получено , 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 количества отсчетов (0 – четное число, 1 – нечетное число). Вычислим гипотетические вероятности 0 и 1 в выходной бинарной последовательности. В таблице 4 приведено количество четных и нечетных отсчетов в гистограммах частот выборок, а также оценки их согласия с равновероятным распределением по критерию .

Таблица 4 – 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 четных отсчетов | 5 118 387 | 1 535 637 |
| *P*-значения статистики (с одной степенью свободы) | 0.313395 | 0.720967 |

Из таблицы видно, что для обеих выборок 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.

**Список литературы**

1. Принципы разработки и модернизации шифровальных (криптографических) средств защиты информации. [Электронный ресурс]. – 2016. – Режим доступа: https://www.tc26.ru. – Дата доступа: 12.04.2018.

2. Крамер, Г. Математические методы статистики / Г. Крамер – М.: Мир, 1976.

3. Зубков, А.М. Об одной статистике для проверки однородности полиномиальных выборок / А.М. Зубков, Б.И. Селиванов // Дискретная математика. – 2014.

4. Csiszar, I. Consistency of the BIC order estimator / I. Csiszar, P. Shields // Electronic research announcments of the American mathematical society. – 1999. – Vol. 5. – P. 123–127.