Лекция №3. Имитация случайных величин и процессов. Базовые датчики

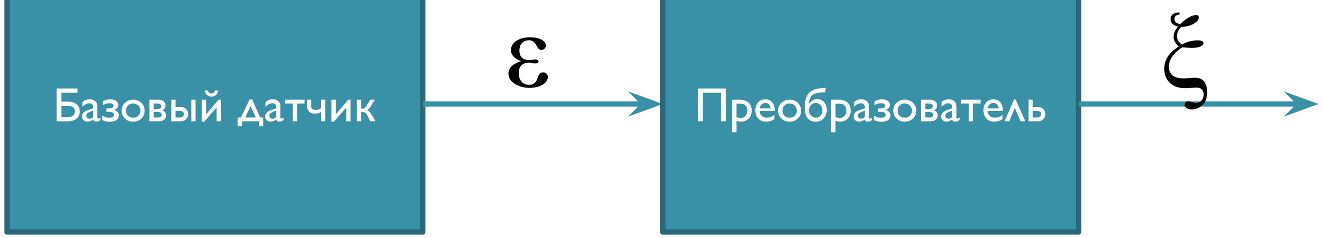
Вопросы

1. Подходы к имитации случайных величин. Понятие базового датчика.
2. Конгруэнтные базовые датчики.
3. Требования к базовым датчикам и их проверка.

1. Подходы к имитации случайных величин. Понятие базового датчика

- Для моделирования влияния неконтролируемых факторов при создании имитационных моделей используют генераторы случайных чисел.
- Предварительно должны проводиться статистические исследования изучаемых признаков, чтобы сделать предположения о параметрах и законе распределения случайных чисел в данном конкретном случае.

Генерация случайных величин



Типы базовых датчиков

По способу генерации случайных величин (СВ) различают:

- Физические базовые датчики
- Таблицы случайных чисел
- Псевдослучайные алгоритмы генерации СВ

Физические базовые датчики

- Используют случайные физические процессы (последние цифры в температуре процессора, при отклике мышки; результаты жеребьевки; данные о поведении элементарных частиц)
- К достоинствам относят их непредсказуемость, в большинстве случаев высокое качество СВ.
- Недостатки – дороговизна и невоспроизводимость сгенерированных СВ.

Физические базовые датчики

- Применяются в основном в научных исследованиях, а также в системах защиты данных.

Таблицы случайных чисел

- Создаются на основе результатов генераций физическими базовыми датчиками.
- Также могут создаваться перед построением модели как результаты натурного экспериментирования, однако при этом они могут не подчиняться равномерному закону.

Таблицы случайных чисел

- К достоинствам относят высокое качество СВ, их воспроизводимость.
- Недостатки – предсказуемость.
- Таблицы широко применяются в различных научных и практических исследованиях.

Псевдослучайные алгоритмы генерации СВ

- К ним относят математические формулы, генерирующие числа, *похожие* на случайные.
- Наиболее распространенные алгоритмы: линейный конгруэнтный метод, мультипликативный конгруэнтный метод, метод Фибоначчи с запаздываниями, регистр сдвига с обратной связью и др.

Псевдослучайные алгоритмы генерации СВ

- Все данные алгоритмы не являются случайными в строгом смысле этого слова.
- К достоинствам относят низкие затраты на генерацию, воспроизводимость СВ.
- Недостатки – предсказуемость, а также специфические недостатки, связанные с низким качеством СВ:
  - Слишком короткий период генерации;
  - Зависимость последующих значений от предыдущих (характерно для линейного конгруэнтного метода с шагом 3);
  - Неравномерность распределения;
  - Обратимость

Псевдослучайные алгоритмы генерации СВ

- Из-за своих недостатков псевдослучайные алгоритмы редко применяются в научных исследованиях, но могут применяться в прикладных исследованиях, в компьютерных играх и др.

Конгруэнтные базовые датчики.

Конгруэнтные базовые датчики

- Конгруэнтные базовые датчики являются одними из простейших и наиболее используемых.
- Тем не менее, качество генерируемых случайных величин данными методами должно проверяться.
- Конгруэнтные базовые датчики генерируют целые числа в интервале от 1 до M-1, где M задается как переменная.
- Также задается первое значение СВ E_0, а последующие генерируются на основании предыдущих чисел

Мультипликативный конгруэнтный базовый датчик

- Выдает целые числа E_i от 0 до М-1
- Каждое последующее рассчитывается на предыдущее по формуле

$$E_i = (\beta * E_{i-1}) \bmod M \qquad (1)$$

где mod – оператор получения остатка от деления

- Для получения чисел в интервале от 0 до 1 Е делится на М

$$\varepsilon_i = E_i / M$$

Мультипликативный конгруэнтный базовый датчик

В формуле (1)

- β – множитель.

Для 64-разрадных чисел возможное значение $\beta = 2^{32} + 3 = 4\,294\,967\,299$

Для 32-разрадных чисел возможное значение $\beta = 2^{16} + 3 = 65\,539$

- Первое значение случайного числа, необходимое для генерации предыдущих, как правило $E_0=\beta$

Мультипликативный конгруэнтный базовый датчик

В предыдущей формуле

- $M-1$ максимальное генерируемое число.

Для 64-разрядных чисел рекоменду-емое значение

$$M = 2^{63} = 9\,223\,372\,036\,854\,775\,808$$

Для 32-разрядных чисел рекомендуемое значение

$$M = 2^{31} = 2\,147\,483\,648$$

Мультипликативный конгруэнтный базовый датчик

- После определенного количества случайных чисел данный базовый датчик начинает генерировать те же числа (зацикливается).
- Период зависит от значений $E_0;\beta;\ M$
- *Для приведенных ранее значений период*
  - *у 64-базового датчика T=2 305 843 009 213 693 952;*
  - *у 32-базового датчика T=536 870 912;*

Мультипликативный конгруэнтный базовый датчик

- Для конгруэнтных методов характерно наличие высокой автокорреляции, например, для приведенного выше 32-разрядного датчика RANDU характерна автокорреляция с шагом 3.
- *Широкое применение данного датчика в первые десятилетия развития компьютерной техники привело к необходимости перестраивать многие модели в различных науках.*

Линейный конгруэнтный базовый датчик

- Выдает целые числа от 0 до М
- Каждое последующее рассчитывается на предыдущее по формуле

$$E_i = (\beta_1 * E_{i-1} + \ldots \beta_p * E_{i-p} + c) \bmod M$$

- Для получения чисел в интервале от 0 до 1 Е делится на М

$$\varepsilon_i = E_i / M$$

Линейный конгруэнтный базовый датчик

Например, возможные значения параметров следующие (датчик Терпугова)

$$E_0 = \beta_1 = 2\,843\,314\,861$$

$$M = 1\,073\,741\,824$$

Требования к базовым датчикам и их проверка.

Требования к базовым датчикам

- Большой отрезок апериодичности
- Равномерность
- Отсутствие автокорреляции
- Соответствие параметров закону распределения

Большой отрезок апериодичности

- Все псевдослучайные базовые датчики с определенного момента начинают выдавать уже выданные данные.
- Отрезок апериодичности T должен быть как можно больше.

$$E_0 \quad E_1 \quad \dots \quad \underbrace{E_L \quad E_{L+1} \quad \dots \quad E_{L+T-1}}_{T} \quad \underbrace{E_{L+T} \quad \dots \quad E_{L+2T-1}}_{T} \quad \dots$$

Равномерность

- Все сгенерированные значения должны быть равномерно распределенными, то есть на любых отрезках равной длины $[E_1 \ldots E_1 + L]$ и $[E_2 \ldots E_2 + L]$ должно быть примерно одинаковое количество сгенерированных случайных чисел.

Проверка на равномерность

- Генерируется N случайных чисел.
- Интервал [0…1] разбивается на k равных интервалов длиной k. Число интервалов может определяться формулой Стерджесса $k=round(1+3.322*lgN)$
- Определяется число значений, попавших в каждый интервал n_i.

Проверка на равномерность

- Рассчитывается критерий χ^2.

$$\chi^2 = \sum_{i=1}^{k} \frac{(N_i - \frac{N}{k})}{\frac{N}{k}}$$

- Полученное значение сравнивается с табличным значением, выбранным согласно уровню значимости α(как правило 0,05) и числу степеней свободы $f=k-1$. Если рассчитанный χ^2 меньше табличного, то можно считать, что распределение достаточно равномерное.

Отсутствие автокорреляции

- Предыдущие значения случайной величины не должны влиять на последующие (то есть после больших значений не должны идти все время большие, или все время маленькие значений).
- Такое влияние не должно происходить ни на следующем шаге, ни через k шагов.

Отсутствие автокорреляции

- Для проверки на отсутствие автокорреляции строят ряд из сгенерированных случайных чисел E, а также этот же ряд со сдвигом на k E*.
- Взаимосвязь между рядами оценивают при помощи методов корреляционно-регрессионного анализа

Проверка на отсутствие методами корреляционного анализа

- Вычисляют линейный коэффициент корреляции:

$$r = \frac{\overline{E * E^*} - \overline{E} * \overline{E^*}}{\sigma_E \sigma_{E^*}} = \frac{12}{N-k} \sum_{i=1}^{N-k} (x_i - 0.5)(x_{i+k} - 0.5)$$

- Значение менее 0,3 отражают отсутствие умеренной или более тесной связи.

Проверка на отсутствие методами корреляционного анализа

- Проверяют значимость связи при помощи t-критерия Стьюдента:

$$t = \frac{r}{\sqrt{1-r^2}}\sqrt{N-k-2}$$

- Значения, меньшие табличных, выбранных согласно уровню значимости α(как правило 0,05) и числу степеней свободы *f=N-k-2*, отражают незначимую связь (на данном уровне значимости α).

Соответствие параметров закону распределения

- Среднее должно быть примерно равно математическому ожиданию

$$\overline{E} = \frac{\sum_{i=1}^{N-k} E_i}{N-k} \approx M(X) = \int_0^1 pXdX = \frac{1}{2}$$

- Дисперсия сгенерированных значений должна примерно равняться дисперсии теоретического распределения.

$$D(E) = \frac{\sum_{i=1}^{N-k} (E_i - \overline{E})^2}{N-k} \approx D(X) = \int_0^1 p(X - M(X))^2 dX = \frac{1}{12}$$

СПАСИБО ЗА ВНИМАНИЕ