Лабораторная работа 3. Моделирование случайных величин с произвольным распределением на основе равномерного распределения

Цель работы

Цель настоящей работы – освоить средства моделирования случайных величин (СВ) с произвольным распределением на основе равномерного распределения. Построить имитационную модель двух потоков, в котором длительность промежутков времени между поступлениями заявок имеет показательный закон с параметрами λ_1, λ_2. Осуществить проверку статистической гипотезы о соблюдении свойства аддитивности пуассоновского потока (сумма пуассоновских потоков есть поток пуассоновский).

Ход работы

1. Ознакомиться со справочными сведениями; сформулировать особенности пуассоновского потока событий; указать связь (дискретного) пуассоновского потока и (непрерывного) показательного распределения.
2. Запрограммировать предложенный алгоритм генерации пуассоновского потока с использованием MatLab или Python.
3. Создать графическую интерпретацию потока событий.
4. Осуществить проверку гипотезы о виде распределения для суммарного потока.
5. Сравнить интенсивности выборочных и теоретических потоков.
6. Составить и представить преподавателю отчет о работе.

Исходные данные: промежуток наблюдения $[T_1, T_2]$, параметр λ. Значения параметра λ должны быть выбраны в зависимости от номера студента в списке группы N, где $T_1 = N$, $T_2 = N + 100$, $\lambda_1 = \frac{N+8}{N+24}$, $\lambda_2 = \frac{N+9}{N+25}$.

Справочные сведения по моделированию случайных величин

1. Поток событий

Под потоком событий в теории вероятностей понимается последовательность событий, происходящих одно за другим в случайные моменты времени. Примерами могут служить: поток вызовов на телефонной станции; поток включений приборов в бытовой электросети; поток заказных писем, поступающих в почтовое отделение; поток сбоев (неисправностей) электронной вычислительной машины; поток выстрелов, направляемых на цель во время обстрела, и т. п. События, образующие поток, в общем случае могут быть различными, но здесь мы будем рассматривать лишь поток однородных событий, различающихся только моментами появления. Такой поток можно изобразить как последовательность точек $t_1, t_2, \dots, t_k, \dots$ на числовой оси, соответствующих моментам появления событий.

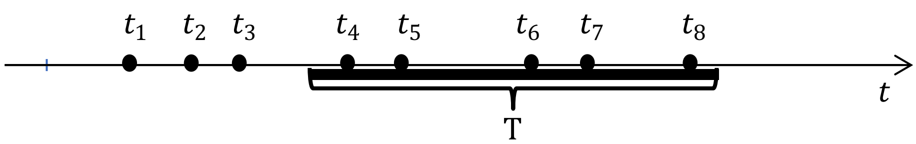


Рисунок 3.1. Поток событий на временной оси

Поток событий называется регулярным, если события следуют одно за другим через строго определенные промежутки времени. Такой поток сравнительно редко встречается в реальных системах, но представляет интерес как предельный случай.

Здесь рассматриваются простейшие потоки со следующими свойствами, определенными содержательно.

1. Поток событий называется стационарным, если вероятность попадания того или иного числа событий на участок времени длиной T зависит только от длины участка и не зависит от того, где именно на оси $[0, t)$ расположен этот участок.

2. Поток событий называется потоком без последействия, если для любых неперекрывающихся участков времени число событий, попадающих на один из них, не зависит от числа событий, попадающих на другие.

3. Поток событий называется ординарным, если вероятность попадания на элементарный участок Δt двух или более событий пренебрежимо мала по сравнению с вероятностью попадания одного события.

Если поток событий обладает всеми тремя свойствами (стационарен, ординарен и не имеет последействия), то он называется *простейшим (или стационарным пуассоновским) потоком*. Название «пуассоновский» связано с тем, что при соблюдении условий 1-3 число событий, попадающих на любой фиксированный интервал времени, будет распределено по закону Пуассона.

Пусть события наступают в случайные моменты времени $t_i = 0, t_1, t_2, \dots, t_k, \dots$, так что $u_k = t_k - t_{k-1}$ $(k \geq 1)$ – интервалы между поступлениями и, кроме того,

$$t_k = u_1 + u_2 + \ldots + u_k.$$

Предполагается, что случайные величины $u_1, u_2, \dots, u_k$ независимы и имеют показательное распределение с параметром λ:

$$\mathrm{P}\{u_k < t\} = 1 - e^{-\lambda t}.$$

Такой входной поток требований в систему является простейшим.

Пусть $\nu(t)$ - число событий, наступивших в интервале времени $(0, t)$. Тогда для простейшего потока справедлива формула:

$$\mathrm{P}\{\upsilon(\mathrm{t}) = \mathrm{k}\} = \frac{(\lambda t)^k}{k!} e^{-\lambda t},$$

означающая, что если длительности промежутков между поступлениями в систему последовательных требований имеют показательный закон, то случайное число событий, наступивших за время t, имеет распределение

Пуассона с параметром λ, а процесс $\nu(t)$ является однородным пуассоновским процессом.

Имеет место и обратное: если число событий $\nu(t)$, наступивших за время t, является процессом Пуассона с интенсивностью λ, то длительности интервалов u_k независимы и имеют одинаковое показательное распределение с параметром λ.

2. Метод обратной функции

Метод обратной функции для формирования СВ с произвольным заданным распределением основан на следующей теореме Смирнова.

Теорема 3.1. (Смирнов Н.В.). Если X удовлетворяет уравнению

$$\int_{-\infty}^{X} F_X(t)dt = \alpha,\ \alpha \in [0;1],$$

где α – СВ, равномерно распределенная на $[0;1]$, то есть $X = F_X^{-1}(\alpha)$, то СВ X распределена по закону $F_X(t)$.

3. Получение СВ, имеющей показательное распределение

Рассмотрим задачу получения СВ с функцией распределения показательного закона распределения.

Решение. Экспоненциальный закон распределения вероятности случайных событий имеет плотность вероятности $f(x)=\lambda e^{-\lambda x}$ при $x\geq0$ и $f(x)=0$ при $x<0$. Функция распределения имеет вид: $F(x)=1-e^{-\lambda x}$.

С помощью датчика случайных чисел получим случайное число, равномерно распределенное в интервале от 0 до 1.

Величины r и F в методе обратной функции полагаются равносильными. Поскольку их значения расположены в одном интервале, то, заменяя F на случайное число r, имеем: $r = 1 - e^{-\lambda x}$.

Используя понятие обратной функции, выражаем искомую величину х из этого выражения и получаем: $x = (-1/\lambda)\cdot\ln(1 - r)$. Так как $(1 - r)$ и r равносильны в смысле их применения в вычислениях, то полагают, что соответствующее значение СВ $x = (-1/\lambda) \cdot \ln(r)$.

4. Алгоритм построения модели пуассоновского потока (согласно теореме Смирнова)

Алгоритм построения модели пуассоновского потока (согласно теореме Смирнова) имеет следующий вид.

1. Генерируем случайное вещественное число ξ_i в диапазоне (0;1).
2. Преобразуем это число в значение СВ с показательным распределением с помощью обратного отображения $u_i = -\dfrac{\ln(\xi_i)}{\lambda}$, используя выражение $\xi_i = 1 - e^{-t}$, где t – промежуток времени, равный в нашем случае интервалу u_i между наступлениями событий.
3. Получаем значение момента наступления события $t_i = T_1 + \sum_{j=0}^{i} u_j$.
4. Генерируем массив значений u_i и t_i, пока $t_i \le T_2$. Полученный массив и есть модель пуассоновского потока.

Графическая интерпретация потока событий представлена на рис. 3.1.

Основным свойством пуассоновского потока, обусловливающим его широкое применение при моделировании, является свойство инвариантности по отношению к операции суммирования: результирующий поток суммы пуассоновских потоков тоже является пуассоновским с суммарной интенсивностью:

$$\lambda = \sum_{i=1}^{n} \lambda_i$$

где n – число пуассоновских потоков, участвующих в суммировании, λ_i – интенсивность i-ого потока.

Предположим, что имеется два точечных процесса и результирующий процесс формируется наложением (суммированием) этих процессов. Тогда результирующий поток соответствует потоку, изображенному на рисунке 3.2.

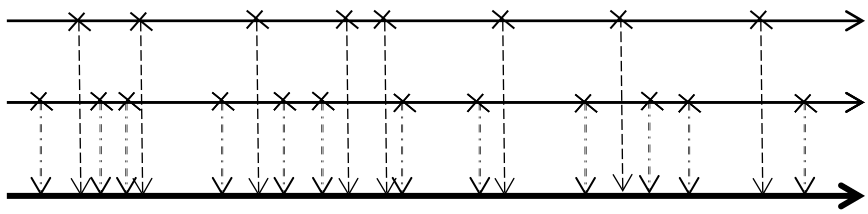


Рисунок 3.2. Графическая иллюстрация суммарного потока на временной оси

5. Алгоритм построения модели суммы потоков

1. Получим выборку объема 50 реализаций каждого из потоков $X_1(t)$ и $X_2(t)$ с интенсивностями λ_1, λ_2. на основе алгоритма 4.
2. Получим поток $X(t)$ с интенсивностью $\lambda = \lambda_1 + \lambda_2$. на основе алгоритма 4.
3. Получим 50 реализаций потока $X_{np}(t)$ как сумму потоков $X_1(t)$ и $X_2(t)$ (Рисунок 3.2)
4. Осуществим статистическую обработку полученных выборок (50 реализаций) для каждого потока $X_{\text{пр}}(t)$, $X_1(t)$, $X_2(t)$, $X(t)$.
5. Произведем расчет выборочных характеристик параметров для каждого из потоков:
6. Сравним полученные значения:

– $\hat{\lambda}$ ($X_{\text{пр}}(t)$) и λ ($X(t)$), $\hat{\lambda}_1$ и λ_1, $\hat{\lambda}_2$ и λ_2 для каждого потока;

– $\hat{\lambda}$ потока X, $\hat{\lambda}$ потока $X_{\text{пр}}$ и $\hat{\lambda}$, полученного сложением $\hat{\lambda}_1$ потока X_1(t) и $\hat{\lambda}_2$ потока X_2(t).

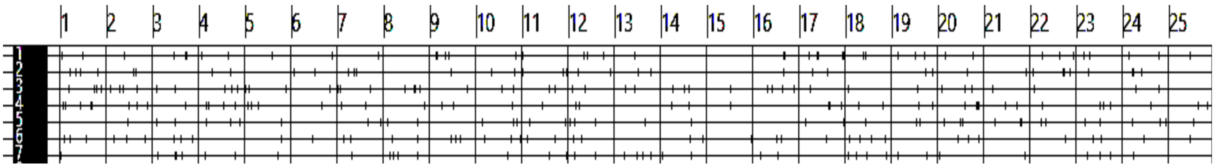
– математическое ожидание $M\eta = \frac{1}{N}\sum_{l=1}^{L} \eta_l \cdot n_l$ и дисперсию потоков $D\eta = \frac{1}{N}\sum_{l=1}^{L}(\eta_l - M\eta)^2 \cdot n_l$.

Текст из лекции-презентации

Алгоритм статистической обработки выборки потоков

1) Генерируем K=50 потоков.

2) Разбиваем интервал $[T_1;T_2]$ на 25 одинаковых промежутков, равных $\Delta t = \frac{T_2 - T_1}{25}$.



3) Формируем вспомогательную таблицу.

	$[T_1, T_1+\Delta t)$	$[T_1+\Delta t, T_1+2\Delta t)$		$[T_1+24\Delta t, T_1+25\Delta t)$
Номер промежутка	1	2	…	25
Итерация 1	$X_1^{(1)}$	$X_1^{(2)}$	…	$X_1^{(25)}$
Итерация 2	$X_2^{(1)}$	$X_2^{(2)}$	..	$X_2^{(25)}$
….	..	..	..	..
Итерация K	$X_K^{(1)}$	..	..	$X_K^{(25)}$

4) Из двумерного массива $\{X_j^{(i)}\}$ выбираем варианты

Параметр	Уникальные значения (варианты)			
	1	2	….	L
η_l (варианты)	η_1	η_2	…..	η_L
n_l (частоты η_l)	n_1	n_2	….	n_L
$\eta_l * n_l$	$\eta_1 * n_1$	$\eta_2 * n_2$	….	$\eta_L * n_L$

$n_l^{\text{теор}}$	$n_1^{\text{теор}}$	$n_2^{\text{теор}}$	….	$n_L^{\text{теор}}$
$\dfrac{(n_l - n_l^{\text{теор}})^2}{n_l^{\text{теор}}}$				

5) Вычисляем выборочные числовые характеристики

$\hat{\lambda} * \Delta t = M\eta = \frac{1}{N}\sum_{l=1}^{L} \eta_l n_l$	Выборочная интенсивность наступления событий $\hat{\lambda}\Delta t$ как выборочное математическое ожидание СВ $\lambda\Delta t$ (число событий, наступивших в ед.времени).
$\widehat{p_l} = \frac{(\hat{\lambda} * \Delta t)^{\eta_l}}{\eta_l!} * e^{-\hat{\lambda} * \Delta t}$	Оценка «теоретической» вероятности по формуле Пуассона
$n_l{}^{\text{теор}} = \widehat{p_l} \cdot N$	Оценки «теоретических» частот вариантов
$N = \sum_{l=1}^{L} n_l$	Объем выборки как сумма частот
$\chi^2{}_{\text{практ}} = \sum_{l=1}^{L} \frac{(n_l - n_l{}^{\text{теор}})^2}{n_l{}^{\text{теор}}}$	Критерий χ^2 (хи-квадрат), проверяет значимость расхождения эмпирических (наблюдаемых) и теоретических (ожидаемых) частот

6) Вычисляем значение квантиля хи квадрат $\chi^2{}_{(\varepsilon,n)}{}^{\text{крит.}}$

Например, уровень значимости ε (вероятность наблюдаемого значения быть случайным отклонением) равен 0,05, а число степеней свободы равно числу карманов (интервалов) за вычетом единицы и числа параметров закона распределения (n=25-1-1=23).

7) Если $\boldsymbol{\chi^2{}_{\text{практ}} < \chi^2{}_{(\varepsilon,n)\ \text{крит}}}$, то гипотеза о «пуассоновости» потока не отвергается (гипотеза согласуется с выборочными данными).

Значение квантиля $\chi^2_{(\varepsilon,n)^{\text{крит}}}$ можно найти с помощью функции `chi2inv` в MatLab и функции `chi2.ppf` из пакета `scipy.stats.distributions` в Python. Вычислить теоретическую функцию плотности вероятности Пуассона можно с помощью функции `poisspdf` в MatLab и `scipy.stats.poisson.pmf` в Python.

Результаты расчетов значений критерия хи-квадрат можно проверить с помощью функции `chi2gof` в MatLab или `scipy.stats.chisquare` в Python.

Получить значение случайной величины, имеющей равномерное распределение в диапазоне [0; 1), можно с помощью функции `rand` в MatLab или функции `random.random` в Python.