ЛАБОРАТОРНА РОБОТА № 3

ГЕНЕРУВАННЯ ДИСКРЕТНИХ ВИПАДКОВИХ ВЕЛИЧИН ІЗ ЗАДАНИМ ЗАКОНОМ РОЗПОДІЛУ

Створення генераторів випадкових чисел з довільним, наперед заданим законом розподілу дуже непроста задача. Тому випадкові числа з необхідним законом розподілу отримують не безпосередньо, а шляхом перетворення випадкових чисел, що мають деякий початковий розподіл. До початкового розподілу висувають такі вимоги: простота отримання чисел на ЕОМ; зручність перетворення випадкових чисел у розподіл із заданим законом. Встановлено, що рівномірний закон розподілу достатньою мірою задовольняє цим вимогам. Отже, моделювання випадкових процесів з довільним законом розподілу може бути побудоване на використанні датчика випадкових чисел, рівномірно розподілених в інтервалі [0;1].

Розглянемо дискретну випадкову величину X, що приймає n значень $x_1, x_2,...,$ x_n з імовірностями $p_1, p_2,..., p_n$. Ця величина задається таблицею розподілу

X	x_1	x_2	…	x_n
P	p_1	p_2	…	p_n

причому $\sum_{i=1}^{n} p_i = 1$.

Для моделювання такої дискретної випадкової величини відрізок [0;1] розбивають на n послідовних відрізків $\Delta_1, \Delta_2,..., \Delta_n$, довжини яких дорівнюють відповідним ймовірностям $p_1, p_2, ..., p_n$, тобто $\Delta_i=p_i$, $i=1,2,\ldots,n$.

За допомогою рандомного генератора отримують випадкову величину $R\{r_1,r_2,\ldots,r_n\}$, рівномірно розподілену в інтервалі [0,1]. Якщо випадкове число r_k, що формується генератором випадкових чисел, котрі відповідають рівномірному закону розподілу на інтервалі [0,1], потрапляє до інтервалу Δ_i, то випадкова величина X набуває значення x_i. з імовірністю p_i. Дійсно:

$$P(l_{i-1} < R \le l_i) = \int_{l_{i-1}}^{l_i} f_R(x)dx = p_i,$$

де $l_{i-1} = \sum_{k=1}^{i-1} p_k$, $l_i = \sum_{k=1}^{i} p_k = l_{i-1} + p_i$, $f_R(x)$ – густина розподілу імовірності випадкової величини R:

$$f_R(x) = \begin{cases} 1, & x \in [0;1] \\ 0, & x \notin [0;1] \end{cases}.$$

Має місце твердження: якщо кожному випадковому числу $r_k \in [0,1]$, яке потрапило в інтервал Δ_i, поставити у відповідність можливе значення x_i з ймовірністю p_i, то величина, яка розігрується, буде мати заданий закон розподілу.

Приклад. Змоделювати 8 значень дискретної випадкової величини X, яка задана таблицею розподілу:

X	x_1=3	x_2=11	x_3=24
P	p_1=0,25	p_2=0,16	p_3=0,59

Розв’язок.

1. Розіб’ємо інтервал [0;1] точками з координатами 0,25; 0,25+0,16=0,41 на три часткових інтервали:

Δ_1=[0;0,25), Δ_2=[0,25;0.41), Δ_3=[0,41;1].

2. Згенеруємо за допомогою комп'ютера 8 випадкових чисел $r_k \in [0,1]$, наприклад, r_1 = 0,10; r_2 = 0,37; r_3 = 0,08; r_4 = 0,99; r_5 = 0,12; r_6 = 0,66; r_7 = 0,31; r_8 = 0,85.

3. Випадкове число r_1 = 0,10 належить першому частковому інтервалу, тому випадкова величина, яка розігрується прийняла можливе значення x_1 = 3. Випадкове число r_2 = 0,37 належить другому частковому інтервалу, тому величина, яка розігрується прийняла можливе значення x_2 = 11. Аналогічно отримаємо решту можливих значення дискретної випадкової величини X.

Результат: послідовність змодельованих можливих значень дискретної випадкової величини X така: {3; 11; 3; 24; 3; 24; 11; 24}.

Перевірка гіпотези про закон розподілу методом гістограм

Нехай у результаті експерименту отримано n значень $x_1, x_2, \ldots, x_n$ випадкової величини X і всі вони лежать у межах $a < x_i < b$.

Суть перевірки по гістограмі така:

1. Інтервал $[a; b]$ розбивається на L підінтервалів довжинами Δ_j. На практиці, як правило, усі підінтервали вибираються однакової довжини. Тоді $\Delta_j=(b-a)/L$, $j=\overline{1,L}$. Число підінтервалів L можна встановити аналітично за формулою Стерджеса: L=1+3,322lg$n$ де n – кількість значень випадкової величини. Нижче наведено таблицю, яка визначає число підінтервалів за об'ємом вибірки n:

n	15-22	23-45	46-90	91-180	181-360	361-710
L	5	6	7	8	9	10

Тоді при генеруванні послідовності $\{x_i\}$ кожне з чисел x_i потрапляє в один з підінтервалів. Усього в кожен j-й підінтервал потрапляє n_j чисел послідовності $\{x_i\}$, $i=\overline{1,n}$, (n_j називають частотою потраплянь випадкових чисел послідовності $\{x_i\}$ у кожний з підінтервалів), причому $n=\sum_{j=1}^{L} n_j$. Відносною частотою потрапляння випадкових чисел послідовності $\{x_i\}$ в кожний з підінтервалів називають величину $w_j=n_j/n$.

2. Над кожним з підінтервальних розбиттів будується прямокутник, площа якого дорівнює частоті потрапляння n_j у цей підінтервал. Висота кожного прямокутника дорівнює частоті n_j, поділеній на Δ_j. Отриману ступінчасту лінію називають *гістограмою*. Таким чином, **гістограма є графічним зображенням залежності частоти потрапляння елементів вибірки від відповідного інтервалу групування.**

Знаходження числових характеристик послідовностей випадкових чисел

Нехай у результаті експерименту отримано скінченну послідовність $\{x_i\}$, i=1,2,…,N змодельованих значень випадкової величини X. Таку скінченну послідовність називають *вибіркою* з випадкового процесу. Для вибірки вводяться поняття *вибіркового математичного сподівання* $\overline{x}_N$ і *вибіркової дисперсії* σ_N^2:

$$\overline{x}=\sum_{i=1}^{n}x_i p_i=\frac{1}{n}\sum_{i=1}^{n}x_i\,,$$

$$\overline{x^2}=\sum_{i=1}^{n}x_i^2 p_i=\frac{1}{n}\sum_{i=1}^{n}x_i^2$$

$$\sigma^2=\sum_{i=1}^{n}(x_i-\overline{x})^2 p_i=\frac{1}{n}\sum_{i=1}^{n}(x_i-\overline{x})^2=\overline{x^2}-(\overline{x})^2\,.$$

Значення $\overline{x}_N$ і σ_N^2, при відносно великих n, можна прийняти в якості оцінок математичного сподівання $M\{X\}$ і дисперсії $D\{X\}$ величини X, тобто, $M\{X\}=\overline{x}\cong\overline{x}_N$, $D\{X\}=\sigma^2\cong\sigma_N^2$. Наближені рівності стають точними при $N\to\infty$. Вибіркове середньоквадратичне відхилення σ_N дорівнює квадратному кореню з вибіркової дисперсії $\sigma_N=\sqrt{\sigma_N^2}$.

Моделювання випадкової величини, розподіленої за законом Пуассона.

Цілочислова випадкова величина X має ***біноміальний закон*** розподілу, якщо ймовірність її можливих значень обчислюється за формулою Бернуллі:

$$P(X=k)=C_n^k p^k q^{n-k},\;\; k=0,1,2,...,n. \qquad (1)$$

Біноміальний розподіл є дискретним розподілом імовірностей з параметрами n і p для кількості успішних результатів у послідовності з n незалежних експериментів, для кожного з яких ставиться питання "*так-ні*". Імовірність виникнення успішного результату для кожного випробування задається параметром p, а імовірність виникнення неуспішного результату, відповідно, дорівнює q=1–p. Випадкова величина, що має біноміальний розподіл коротко записується так: $X \in B\,(n,p)$

Числові характеристики *біноміального розподілу*: **математичне сподівання $M(X)$ = np, дисперсія $D(X)$ = npq.**

Розподілом Пуассона називається закон розподілу дискретної випадкової величини *X*, яка може приймати будь-які цілі невід'ємні значення *k* в замкнутому проміжку [0, *n*], що описується формулою Пуассона:

$$P(X=k)=\frac{\lambda^k}{k!}e^{-\lambda}, \qquad (2)$$

при цьому λ=const – параметр розподілу Пуассона. Дана формула визначає ймовірності подій у схемі повторних незалежних випробувань Бернуллі (1) коли $p\to 0$, $n\to\infty$, $\lambda=np$. Закон Пуассона описує число подій, що відбуваються за однакові проміжки часу, за умови, що ці події відбуваються незалежно одна від одної. Розподілом Пуассона добре описуються число викликів на телефонну станцію за певний час доби, число самородків при розробці родовищ золота й ін. Розподіл Пуассона відіграє ключову роль у теорії масового обслуговування. Випадкова величина, що має розподіл Пуассона коротко записується так: $X\in P(\lambda)$.

Числові характеристики *розподілу Пуассона*: **математичне сподівання** $M(X)=\lambda$**, дисперсія** $D(X)=\lambda$**.**

Як зазначалося вище, закон Пуассона є граничним для біноміального розподілу. Це гранична властивість біноміального розподілу часто знаходить застосування на практиці. Припустимо, що проводиться велика кількість *n* незалежних дослідів, у кожному з яких подія *A* має дуже малу ймовірність *p*. Тоді для обчислення ймовірності того, що подія *A* відбудеться рівно *k* разів, замість (1) можна скористатися формулою (2), як наближеною, де $\lambda=np$. На підставі цього, процедура моделювання випадкової величини *X*, розподіленої за законом Пуассона із заданим параметром λ, полягає у проведенні $n=\frac{\lambda}{p}$ випробувань, де ймовірність появи кожної події є малою величиною: $p\le 0{,}1$ (рідкісна подія).

Алгоритм моделювання такий:

1. Визначають число випробувань $n=\frac{\lambda}{p}$, де $p\le 0{,}1$. Лічильнику числа подій *k* присвоюється значення 0 (*k*:=0).

2. За допомогою датчика випадкових чисел $\{r_i\}$ з рівномірним розподілом в інтервалі [0,1] вибирають число r_i і перевіряють умову $r_i\le p$. Якщо ця умова

виконується, то k збільшується на одиницю (k:=k+1), якщо умова не виконується – k не змінюється.

3. Після проведення n таких випробувань вміст лічильника числа подій k використовується як випадкове число із законом розподілу Пуассона з параметром λ.

4. Пункти 2 і 3 повторюють, поки не буде досягнуто потрібного об'єму вибірки.

Порядок виконання роботи

Частина 1

1. Змоделювати послідовність із N=100 значень дискретної випадкової величини X, заданої одним із варіантів у таблиці 1. Номер варіанту дорівнює порядковому номеру студента у списку академічної групи.

Таблиця 1 – Таблиця розподілів

Варіант	Таблиця розподілів							
1	x_i	5	7	17	19	21	25	55
	p_i	0.01	0.05	0.3	0.3	0.3	0.02	0.02
2	x_i	1	3	7	10	15	18	23
	p_i	0.1	0.05	0.02	0.05	0.25	0.33	0.2
3	x_i	2	3	5	12	21	33	44
	p_i	0.1	0.15	0.2	0.05	0.02	0.33	0.15
4	x_i	5	8	13	16	21	24	29
	p_i	0.1	0.02	0.25	0.15	0.35	0.03	0.1
5	x_i	2	3	5	8	11	15	20
	p_i	0.1	0.15	0.25	0.05	0.05	0.3	0.1
6	x_i	1	8	17	23	37	42	50
	p_i	0.01	0.15	0.05	0.25	0.5	0.02	0.02
7	x_i	1	4	12	16	25	33	37

	p_i	0.05	0.25	0.25	0.15	0.13	0.1	0.07
8	x_i	1	10	15	23	29	38	42
	p_i	0.02	0.05	0.1	0.28	0.23	0.22	0.1
9	x_i	2	3	7	12	19	23	30
	p_i	0.04	0.15	0.2	0.25	0.2	0.15	0.01
10	x_i	1	5	7	14	21	26	31
	p_i	0.34	0.28	0.16	0.15	0.05	0.01	0.01
11	x_i	3	5	8	14	27	29	35
	p_i	0.02	0.07	0.1	0.19	0.19	0.2	0.23
12	x_i	7	16	28	33	39	46	56
	p_i	0.01	0.05	0.07	0.1	0.17	0.25	0.35
13	x_i	5	6	8	13	19	26	36
	p_i	0.05	0.07	0.2	0.23	0.17	0.23	0.05
14	x_i	3	9	18	23	29	27	45
	p_i	0.05	0.14	0.2	0.22	0.17	0.14	0.08
15	x_i	13	16	28	33	39	47	52
	p_i	0.08	0.14	0.25	0.16	0.25	0.09	0.03
16	x_i	1	6	8	13	19	24	27
	p_i	0.09	0.1	0.21	0.17	0.23	0.15	0.05
17	x_i	4	6	10	14	16	20	24
	p_i	0.04	0.1	0.1	0.27	0.33	0.13	0.03
18	x_i	2	6	12	16	22	26	32
	p_i	0.02	0.14	0.24	0.27	0.2	0.1	0.03
19	x_i	3	6	9	13	19	27	31
	p_i	0.04	0.12	0.22	0.28	0.2	0.1	0.04
20	x_i	1	3	8	11	19	29	33

	p_i	0.02	0.26	0.18	0.32	0.16	0.02	0.04

2. Визначити вибіркове математичне сподівання та вибіркову дисперсію отриманої дискретної випадкової величини та порівняти їх з теоретичними значеннями.

3. Побудувати частотну таблицю 2 (кількість інтервалів не менше 10), вивести її на екран.

Таблиця 2 – Частотна таблиця

Інтервал	Частота потрапляння	Відносна частота потрапляння
Δ_1	n_1	w_1
Δ_2	n_2	w_2
…	…	…
Δ_L	n_L	w_L

4. Побудувати гістограму та оцінити за її допомогою закон розподілу випадкової величини X.

5. Повторити виконання роботи для N=1000. Порівняти результати.

Частина 2

1. Змоделювати послідовність із N=100 значень випадкової величини X, розподіленої за законом Пуассона.

2. Визначити вибіркове середнє і вибіркову дисперсію та порівняти їх з теоретичними значеннями.

3. Побудувати гістограму та оцінити за її допомогою закон розподілу випадкової величини X.

4. Повторити п.1-3 для N=1000. Порівняти результати.

Вибір параметрів моделювання: n – порядковий номер студента у журналі, для моделювання розподілу Пуассона $\lambda=n$, $p=0{,}1$.

Контрольні питання

1. Яких значень може набувати випадкова величина з пуассонівським розподілом?
2. Запишіть вираз для розподілу Пуассона.
3. Чому дорівнює математичне сподівання пуассонівської випадкової величини?
4. Чому дорівнює дисперсія пуассонівської випадкової величини?