Лабораторна робота

Закони розподілу ймовірностей дискретних випадкових величин

Дискретні випадкові величини

Табличний запис закону розподілу — це таблиця значень x_k випадкової величини та відповідних їхніх ймовірностей p_k:

$\boldsymbol{x_k}$	x_1	x_2	…	x_n
$\boldsymbol{p_k}$	p_1	p_2	…	p_n

За допомогою табличного запису закону розподілу можна визначити функцію розподілу $F(x)$ випадкової величини X за формулою:

$$F(x) = P\{X < x\} = \sum_{k:\, x_k < x} p_k \ ,$$

у якій сумування проводиться за всіма індексами k, для яких $x_k < x$.

Реалізація в R:

```
# Функція розподілу
pdiscrete <- function(x, xk, pk) {
  ind <-  which(xk < x)
  return (sum(pk[ind]))
}
# Табличний запис закону розподілу
xk <- c(-2, 1, 4, 6)
pk <- c(0.2, 0.1, 0.3, 0.4)
# Ймовірність того, що значення випадкової величини < 5
p <- pdiscrete(5,xk,pk)

##########################################################

# Спосіб з використанням функцій cumsum() та stepfun() в R
x <- xk; y <- c(0, cumsum(pk))
# функція cumsum() повертає вектор, елементи якого є накопичувальними сумами елементів аргументу

myFun <- stepfun(x, y, f=1, right=TRUE)
# Побудова графіка функції розподілу
plot.stepfun(myFun, xlab="x", ylab="F(x)",
             main = "Функція розподілу дискретної випадковох величини", verticals=FALSE, pch=16)
```

Біномний розподіл

Нехай проводиться серія з n незалежних випробувань, кожне з яких закінчується або успіхом, або неуспіхом. Нехай у кожному випробуванні ймовірність успіху p, а ймовірність неуспіху $q = 1 - p$. З таким випробуванням можна пов'язати випадкову величину X – число успіхів у серії з n випробувань. Ця величина приймає цілі значення від 0 до n. Її розподіл називається біномним (біноміальним) і визначається формулою Бернуллі:

$$p_k = P\{X = k\} = C_n^k p^k q^{n-k},$$

де $k = 0, 1, \dots, n$.

Легко переконатися, що $\sum_{k=0}^{n} p_k = 1$.

В R для обчислення ймовірності і функції розподілу випадкової величини, що має біномний розподіл, призначені функції $\boldsymbol{dbinom(k, n, p)}$ і $\boldsymbol{pbinom(k, n, p)}$, значення яких p_k і $F(k) = P\{X \le k\}$ відповідно.

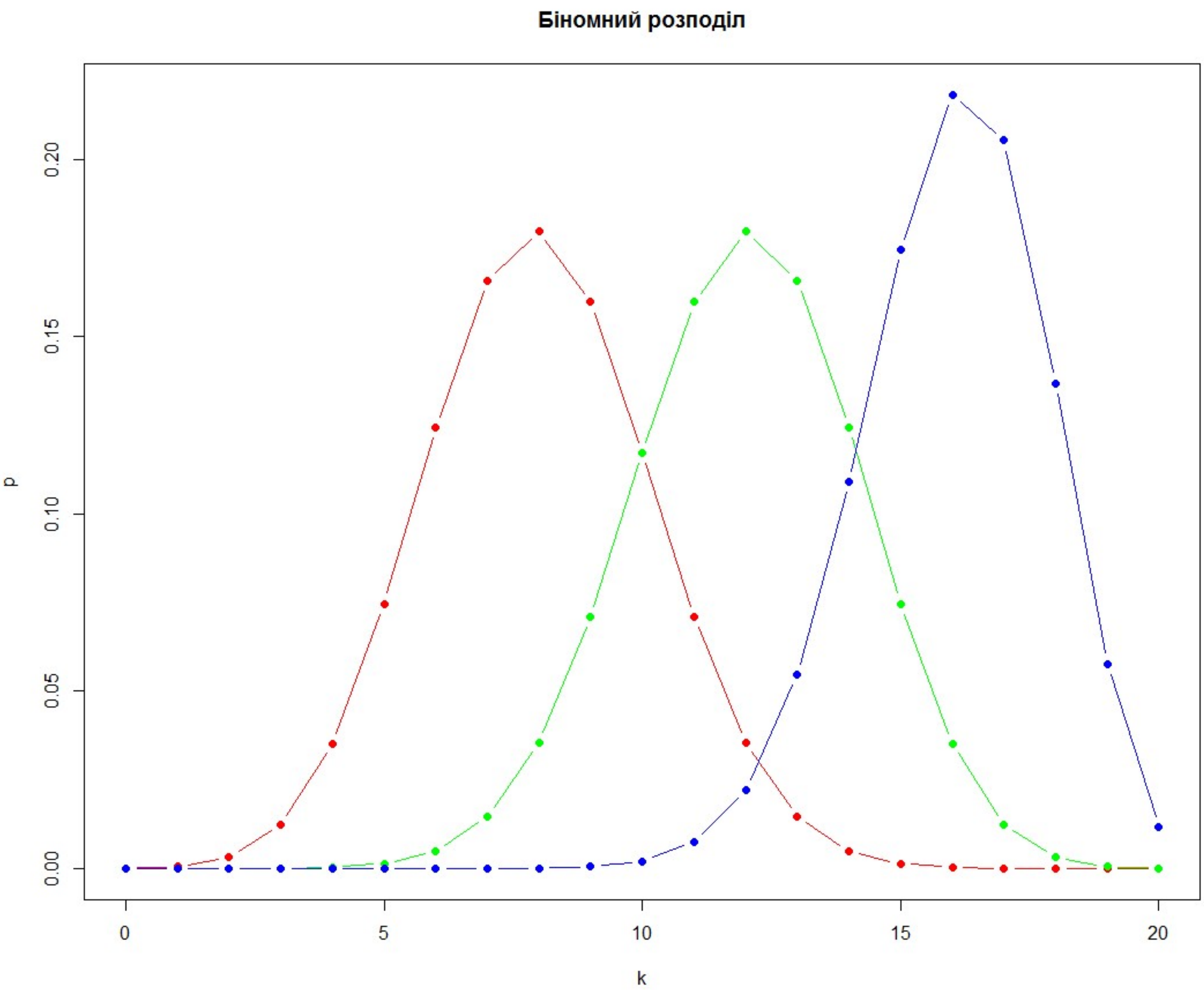
Приклад. Побудувати біномний розподіл для серії з 20 незалежних випробувань з ймовірністю успіху 0.4, 0.6 і 0.8. Побудувати графік розподілу. Для $p = 0.4$ знайти значення k, для якого величина $P\{X = k\}$ максимальна. Перевірити рівність $\sum_{k=0}^{n} p_k = 1$. Знайти ймовірність попадання значень випадкової величини в проміжок [2, 5].

Зразок виконання в R:

```
k <- 0:20
P4 <- dbinom(k,20,0.4)
P6 <- dbinom(k,20,0.6)
P8 <- dbinom(k,20,0.8)

# Графік розподілу
plot(k, P4, ylim=range(c(P4,P6,P8)), xlim=range(k), col="red", type="b",
     pch=16, xlab="k", ylab="p", main = "Біномний розподіл")
lines(k, P6, col="green", type="b", pch=16)
lines(k, P8, col="blue", type="b", pch=16)

# Обчислення найвірогіднішого k та відповідної ймовірності
kp <- k[which(P4 == max(P4))]
maxP <- max(P4)

# Перевірка (=1)
result <- sum(P4)==1

# Ймовірність попадання значень випадкової величини в проміжок [2,5]
p <- pbinom(5,20,0.4) - pbinom(1,20,0.4)
```

Результат:

k	int [1:21] 0 1 2 3 4 5 6 7 8 9 ...
kp	8L
maxP	0.179705787754689
p	0.125074923346629
P4	num [1:21] 3.66e-05 4.87e-04 3.09e-03 1.…
P6	num [1:21] 1.10e-08 3.30e-07 4.70e-06 4.…
P8	num [1:21] 1.05e-14 8.39e-13 3.19e-11 7.…
result	TRUE



Геометричний розподіл

Зі схемою випробувань Бернуллі можна пов’язати ще одну випадкову величину: X – число випробувань до першого успіху. Ця величина приймає значення від 0 до $+\infty$, і її розподіл визначається формулою

$$p_k = P\{X = k\} = q^k p,$$

де $k = 0, 1, \dots,\ 0 < p < 1, q = 1 - p$.

Як і для біномного розподілу, $\sum_{k=0}^{\infty} p_k = 1$. В R для обчислення ймовірності і функції розподілу випадкової величини, що має геометричний розподіл, призначені функції $\boldsymbol{dgeom(k,p)}$ і $\boldsymbol{pgeom(k,p)}$, значення яких p_k і $F(k) = P\{X \le k\}$ відповідно.

Гіпергеометричний розподіл

Нехай в партії з N виробів наявні M $(M < N)$ якісних і $(N - M)$ дефектних виробів. Якщо випадковим способом зі всієї партії вибрати контрольну партію з n виробів, то число якісних виробів в цій партії буде випадковою величиною, яку позначимо X. Її розподіл має вигляд:

$$p_k = P\{X = k\} = \frac{C_M^k C_{N-M}^{n-k}}{C_N^n}, k = 0, 1, \dots, \min(n, M),$$

і називається гіпергеометричним.

Для будь-яких значень параметрів, що входять в розподіл $\sum_{k=0}^{\min(n,M)} p_k = 1$.

В R для обчислення ймовірності і функції розподілу випадкової величини, що має гіпергеометричний розподіл, призначені функції $\boldsymbol{dhyper(k, M, N - M, n)}$ і $\boldsymbol{phyper(k, M, N - M, n)}$, значення яких p_k і $F(k) = P\{X \le k\}$ відповідно.

Розподіл Пуассона

Розподіл Пуассона має випадкова величина X, що приймає значення $k = 0,1,2, \dots$ з ймовірностями

$$p_k = P\{X = k\} = \frac{\lambda^k}{k!} e^{-\lambda}$$

$$p_k = P(\mu = k) = \frac{\lambda^k}{k!} e^{-\lambda}$$ k=0,1,2,…

де $\lambda > 0$ – параметр розподілу Пуассона. При будь-якому $\lambda > 0, \sum_{k=0}^{\infty} p_k = 1$.

В R для обчислення ймовірності і функції розподілу випадкової величини, що має розподіл Пуассона, призначені функції $\boldsymbol{dpois}(\boldsymbol{k}, \boldsymbol{\lambda})$ і $\boldsymbol{ppois}(\boldsymbol{k}, \boldsymbol{\lambda})$, значення яких p_k і $F(k) = P\{X \leq k\}$ відповідно.

У базовій версії R є функції для роботи з поширеними законами розподілу ймовірностей. Залежно від призначення, імена цих функцій починаються з однієї з наступних чотирьох букв:

- d (від "*d ensity*", щільність): функції щільності ймовірності ("функція розподілу мас" для дискретних величин);
- p (від "*p robability*", ймовірність): кумулятивні функції розподілу ймовірностей;
- q (від "*q uantile*", квантиль): функції для знаходження квантилів того чи іншого розподілу;
- r (від "*r andom*", випадковий): функції для генерації випадкових чисел відповідно до параметрів того чи іншого закону розподілу ймовірностей.

Щоб переглянути перелік розподілів для яких реалізовані всі ці функцій в R необхідно в консолі набрати **help(Distributions)** або **?Distributions**.

Індивідульні завдання

Завдання 1

Побудувати функцію розподілу та графік функції розподілу для випадкової величини з розподілом згідно свого варіанту.

№ 1

x	0	0,1	0,2	0,3	0,4	0,5
p	0,3	0,1	0,2	0,15	0,2	0,05

№ 2

x	3	3,2	3,3	3,6	3,9	4
p	0,1	0,11	0,2	0,16	0,09	0,34

№ 3

x	0,3	0,7	1,1	1,4	1,95	2,3
p	0,33	0,05	0,25	0,15	0,12	0,1

№ 4

x	0,7	0,9	1,15	1,32	1,56	1,7
p	0,21	0,1	0,1	0,16	0,19	0,24

№ 5

x	5	6	6,6	7	7,7	8
p	0,3	0,15	0,15	0,1	0,25	0,05

№ 6

x	2,4	3,2	3,4	3,6	3,9	7
p	0,15	0,1	0,1	0,16	0,1	0,39

№ 7

x	0,35	0,45	0,55	0,7	0,75	0,9
p	0,05	0,15	0,25	0,25	0,12	0,18

№ 8

x	4,7	5	5,1	5,35	5,5	5,7
p	0,3	0,2	0,1	0,1	0,1	0,2

№ 9

x	9,3	9,9	10,5	11,1	11,7	12,3
p	0,05	0,15	0,25	0,19	0,14	0,22

№ 10

x	7,7	8,9	10,1	11	12	13,7
p	0,22	0,1	0,15	0,15	0,1	0,28

№ 11

x	1	2	3,5	5	5,5	6,9
p	0,23	0,1	0,12	0,15	0,25	0,15

№ 12

x	0	0,1	0,2	0,3	0,4	0,5
p	0,2	0,15	0,2	0,2	0,2	0,05

№ 13

x	3	3,2	3,3	3,6	3,9	4
p	0,12	0,11	0,22	0,16	0,05	0,34

№ 14

x	0,3	0,7	1,1	1,4	1,95	2,3
p	0,3	0,05	0,2	0,15	0,12	0,18

№ 15

x	0,7	0,9	1,15	1,32	1,56	1,7
p	0,21	0,2	0,1	0,16	0,15	0,18

№ 16

x	5	6	6,6	7	7,7	8
p	0,2	0,25	0,15	0,1	0,25	0,05

№ 17

x	2,4	3,2	3,4	3,6	3,9	7
p	0,1	0,1	0,15	0,15	0,1	0,4

№ 18

x	0,35	0,45	0,55	0,7	0,75	0,9
p	0,15	0,15	0,15	0,25	0,12	0,18

№ 19

x	4,7	5	5,1	5,35	5,5	5,7
p	0,2	0,25	0,1	0,15	0,1	0,2

№ 20

x	9,3	9,9	10,5	11,1	11,7	12,3
p	0,15	0,05	0,25	0,19	0,14	0,22

№ 21

x	7,7	8,9	10,1	11	12	13,7
p	0,22	0,15	0,1	0,23	0,1	0,2

№ 22

x	1	2	3,5	5	5,5	6,9
p	0,2	0,1	0,1	0,15	0,3	0,15

№ 23

x	0	0,1	0,2	0,3	0,4	0,5
p	0,2	0,15	0,25	0,15	0,2	0,05

№ 24

x	3	3,2	3,3	3,6	3,9	4
p	0,1	0,11	0,25	0,18	0,09	0,27

Завдання 2

Для заданих значень параметрів, згідно свого варіанту, обчислити і побудувати **біномний розподіл** для серії з n незалежних випробувань з ймовірністю успіху p, **розподіл Пуассона** з параметром λ, **геометричний розподіл** з параметрами n, p, **гіпергеометричний розподіл** з параметрами N, M, n.

Для кожного розподілу виконати наступне:

- Перевірити виконання рівності $\sum_k p_k = 1$, де $p_k = P\{X = k\}$;
- Знайти значення k для якого величина $P\{X = k\}$ максимальна;
- Побудувати графік розподілу;

  Графік розподілу – це ломана лінія, вершини якої точки (k, p_k), де k – значення випадкової величини, а p_k– ймовірність цього значення.

- Обчислити ймовірність попадання значень випадкової величини в проміжок $[a, b]$.

Варіант	**n**	**p**	**λ**	**N**	**M**	**a**	**b**
1	20	0,1	1	100	90	2	4
2	22	0,11	0,95	110	100	3	5
3	24	0,12	0,9	120	100	2	5
4	26	0,13	0,85	130	120	3	6
5	28	0,14	0,8	140	120	2	5
6	30	0,15	0,75	150	120	3	6
7	21	0,16	0,7	160	150	2	8
8	23	0,17	0,65	170	150	3	9
9	25	0,18	0,6	170	160	2	7
10	27	0,19	0,55	170	130	3	8
11	29	0,2	0,5	170	165	3	10
12	31	0,21	1,05	105	100	4	10
13	20	0,22	1,1	115	100	3	5
14	22	0,23	1,15	125	100	4	6
15	24	0,24	1,25	135	110	3	6
16	26	0,25	1,3	145	120	4	7