Course: R Language in Computational Probability and Statistics

Lecture 10: Discrete densities and distributions

Lecturer: Nataliia Kruhlova

Генератори псевдовипадкових чисел. Моделювання випадкових величин

В R реалізовано наступні генератори псевдовипадкових чисел:

- лінійний конгруентний;
- генератор WELL;
- генератор Мерсенна-Твістера;
- генератор Вічманна-Хілла;
- Марсалья-Мультикаррі;
- Super-Duper;
- генератор Кнута TAOCP;
- генератор Кнута TAOCP-2002;
- генератор Лек'є-CMRG;
- SFMT.

В R можна також генерувати псевдовипадкові послідовності:

- Холтона;
- Торус;
- Соболя.

По умовчанню генерація відбувається за допомогою генератора **Mersenne-Twister**. В R реалізована можливість змінити базовий датчик.

Функція **set.generator** з пакету **randtoolbox** змінює датчик випадкових чисел на один з наступних датчиків: лінійний конгруентний, Мерсенн-Твістер або WELL. Додатково можна задати і спеціальні параметри кожного з цих датчиків.

В базовій версії **R** для цих же цілей використовується функція **RNGkind().** В якості параметра вибирається назва датчика випадкових чисел: **"Wichmann-Hill","Marsaglia-Multicarry", "Super-Duper", "L'Ecuyer-CMRG ", "Knuth-TAOCP", "Knuth-TAOCP-2002"**.

Також можна безпосередньо генерувати вибірку з рівномірного розподілу однією з наступних функцій:

а) **SFMT**;
б) **halton**;
в) **sobol;**
г) **torus.**

Генератор випадкових чисел (ГВЧ) починає свою роботу з певної точки в просторі можливих чисел. Ця точка називається початковим числом (*seed*).

Можливо зафіксувати це число так, що при повторній генерації буде генеруватися така сама послідовність чисел, що і в перший раз.

Для цього є функція **set.seed().** Ця функція фіксує початкову точку для запуску алгоритму генерування випадкових чисел. В якості аргументу функції вказується випадкове ціле число.

Закони ймовірнісних розподілів, що реалізовані в R

В базовій версії R реалізовані наступні закони розподілу:

- Бета-розподіл (beta)
- Біноміальний розподіл (включно з розподілом Бернуллі) (binom)
- Розподіл Коші (cauchy)
- Розподіл хі-квадрат (chisq)
- Експоненціальний розподіл (exp)
- Розподіл Фішера (f)
- Гама-розподіл (gamma)
- Геометричний розподіл (geom)
- Гіпергеометричний розподіл (hyper)
- Логнормальний розподіл (lnorm)
- Поліноміальний розподіл (multinom)
- Від'ємний біноміальний розподіл (nbinom)
- Нормальний розподіл (norm)
- Розподіл Пуассона (pois)
- Розподіл Стьюдента (t)
- Рівномірний розподіл (unif)
- Розподіл Вейбула (weibull)

Для кожного з цих розподілів реалізовано 4 групи функцій різного призначення. Імена цих функцій починаються однією з чотирьох букв, в залежності від призначення:
° d ("*density*") – щільність ймовірності ("функція розподілу ваги для дискретних величин);
° p ("*probability*") - кумулятивна функція розподілу;
° q ("*quantile*") – функція для обчислення квантілів розподілу;
° r ("*random*") - функція для генерації випадкових величин з відповідним законом розподілу.
Зауважимо, що функція розподілу в R задається у відповідності до західної школи ймовірностей: $F_{\xi}(x) = P\{\xi \le x\}$.

Далі ми розглянемо лише найбільш поширені ймовірнісні закони розподілу.

Розподіл Бернуллі

Дискретна випадкова величина має розподіл Бернуллі, якщо вона приймає лише два значення: 0 (невдача) або 1(успіх).

Розподіл можна задати таблицею:

X	0	1
P	1-p	p

Є частинним випадком біноміального розподілу для $n = 1$.

Приклад. Змоделюйте експеримент 200 підкидань несиметричної монети з ймовірністю випадіння герба 0.7. Знайдіть відносну частоту випадінь герба.

```
> x<-sample(c(0,1),200,replace=T,prob=c(0.3,0.7))
> length(x[x==1])/length(x)
[1] 0.645
```

Біноміальний розподіл

Ймовірність k успіхів в серії незалежних випробувань визначається за формулою:

$$P\{\xi = k\} = C_n^k p^k (1-p)^{n-k}, k = \overline{0,n}.$$

Основними аргументами функцій **dbinom**, **pbinom**, **qbinom**, **rbinom** є:

• **x** — невід'ємний вектор з цілими значеннями (значення випадкової величини).

• **q** — невід'ємний вектор (квантілі).

• **p** — вектор ймовірностей.

• **n** — об'єм вибірки.

• **size** — кількість випробувань Бернуллі.

• **prob** — ймовірність успіху в одному експерименті.

• **lower.tail** — логічний параметр. Якщо його значення TRUE, то використовується $P\{\xi \le x\}$, в іншому випадку використовується $P\{\xi > x\}$.

Приклад. Нехай провели 15 незалежних випробувань з ймовірністю успіху 0.25.

Знайдемо ймовірності P{ξ = k} рівно k успіхів (k = 0,…, 15) в 15 випробуваннях.

```
> dbinom(0:15,15,0.25)
 [1] 1.336346e-02 6.681731e-02 1.559070e-01 2.251991e-01 2.251991e-01
 [6] 1.651460e-01 9.174777e-02 3.932047e-02 1.310682e-02 3.398065e-03
[11] 6.796131e-04 1.029717e-04 1.144130e-05 8.800998e-07 4.190952e-08
[16] 9.313226e-10
```

Приклад. Для розподілу з попереднього прикладу знайдемо значення функції розподілу.

```
> pbinom(seq(-0.5,15.5,by=1),15,0.25)
 [1] 0.00000000 0.01336346 0.08018077 0.23608781 0.46128688 0.68648594
 [7] 0.85163192 0.94337969 0.98270016 0.99580699 0.99920505 0.99988466
[13] 0.99998764 0.99999908 0.99999996 1.00000000 1.00000000
```

Приклад. Нехай провели 15 незалежних випробувань з ймовірністю успіху 0.25. Знайдіть квантілі.

```
> qbinom(seq(0,1,by=0.01),15,0.25)
  [1]  0  0  1  1  1  1  1  1  1  2  2  2  2  2  2  2  2  2  2  2  2
 [22]  2  2  2  3  3  3  3  3  3  3  3  3  3  3  3  3  3  3  3  3  3
 [43]  3  3  3  3  3  4  4  4  4  4  4  4  4  4  4  4  4  4  4  4  4
 [64]  4  4  4  4  4  4  5  5  5  5  5  5  5  5  5  5  5  5  5  5  5
 [85]  5  5  6  6  6  6  6  6  6  6  6  7  7  7  7  8 15
```

Приклад. Згенеруємо вибірку об'ємом 10 елементів з вказаним вище розподілом.

```
> rbinom(10,15,0.25)
 [1] 2 5 2 4 3 4 4 5 3 2
```

Розподіл Пуассона

Дискретна випадкова величина розподілена за законом Пуассона з параметром λ, якщо

$$P\{\xi = k\} = \frac{e^{-\lambda}\lambda^k}{k!}, k \in 0,1,\ldots$$

Функції, які відповідають розподілу:

-dpois(x, lambda, log = FALSE),

-ppois(q, lambda, lower.tail = TRUE, log.p = FALSE),

-qpois(p, lambda, lower.tail = TRUE, log.p = FALSE),

- rpois(n, lambda).

Параметри функцій:

- **x** — невід'ємний цілий вектор,
- **q** – невід'ємний вектор значень функції розподілу,
- **lambda** — невід'ємний аргумент (параметр розподілу).

Приклад. Нехай нам потрібно знайти ймовірність $P\{4 < \xi \le 15\}$, де випадкова величина задана за законом Пуассона з параметром 1.5.

Перший спосіб через суму:

```
> sum(dpois(5:15,1.5))
[1] 0.01857594
```

Другий спосіб через функцію розподілу:

```
> ppois(15,1.5)-ppois(4,1.5)
[1] 0.01857594
```

Геометричний розподіл

Дискретна випадкова величина описується геометричним законом розподілу, якщо ймовірність того, що успіх настане в k-тому випробуванні, обчислюється за формулою:

$$P\{\xi = k\} = (1-\mathrm{p})^{k-1}\mathrm{p}, k \in 1, 2, \ldots$$

Функції, що відповідають даному розподілу:

-dgeom(x, prob, log = FALSE)

-pgeom(q, prob, lower.tail = TRUE, log.p = FALSE)

- qgeom(p, prob, lower.tail = TRUE, log.p = FALSE)

- rgeom(n, prob)

Аргументами цих функцій є:

prob — ймовірність «успіху»

p, x — невід'ємні цілі вектори.

Приклад. Микола вирішив зі своїм другом Валерою зіграти в наступну гру:

він підкидає гральний кубик до появи «шістки», але не більше 10 разів. Якщо «шістка» випадає на i-тому підкиданні, то Валера дає йому 2^i гривень, якщо «шістка» не випаде зовсім, то Микола віддає своєму другу 200 гривень. Знайдіть математичне сподівання виграшу Миколи.

Складемо таблицю розподілу випадкової величини – виграшу:

X	-200	2	4	8	16	32	64	128	256	512	1024
p	0.162	0.167	0.139	0.116	0.096	0.08	0.067	0.056	0.047	0.039	0.032

```
> x<-c(-200,2^(1:10))
> p<-c((5/6)^10,(5/6)^(0:9)/6)
> p
 [1] 0.16150558 0.16666667 0.13888889 0.11574074 0.09645062 0.08037551
 [7] 0.06697960 0.05581633 0.04651361 0.03876134 0.03230112
> sum(x*p)
[1] 49.88974
```

Гіпергеометричний розподіл

Гіпергеометричний розподіл краще всього описується наступною задачею.

Нехай в нас є сукупність із N елементів, в якій є підмножина із M елементів з певною властивістю. Із сукупності ми вибираємо n $(n \leq N)$ елементів. Випадкова величина X – це кількість елементів із певною властивістю серед відібраних елементів. Знаходиться за формулою:

$$P\{\xi = k\} = \frac{C_M^k C_{N-M}^{n-k}}{C_N^n}, k \in \max\{0, n + M - N\}, 1, \dots, \min\{\mathrm{n}, \mathrm{M}\}.$$

Приклад. У Миколи було 20 яблук, серед яких було 5 з черв'яками. До Миколи підійшов Валера і попросив дати йому 3 яблука. Знайдіть математичне сподівання і дисперсію випадкової величини – числа яблук з черв’яками серед тих яблук, що Микола віддав Валері. Вважайте, що яблука візуально всі однакові, тому Микола не може спеціально обрати червиве яблуко.

Складемо таблицю розподілу випадкової величини.

X	0	1	2	3
p	0,3991	0,4605	0,1316	0,0088

```
> p<-dhyper(0:3,5,15,3)
> p
[1] 0.39912281 0.46052632 0.13157895 0.00877193
> x<-0:3
> ex<-sum(x*p)
> ex
[1] 0.75
> v<-sum((x-ex)^2*p)
> v
[1] 0.5032895
```