Разпределения и R

Дискретни разпределения

Бернулиево разпределение

Бернулиевото разпределение е дискретно разпределение и описва опит, при който вероятността за успех е p, а за неуспех q = 1 - p.
Ако успех означим с 1, а провал с 0, то:

$$P(X = 0) = 1 - p, P(X = 1) = p$$

Очакване и дисперсия:

$$EX = p, VarX = pq$$

Биномно разпределение

Биномното разпределение е дискретно разпределение на броя успехи при проведени n независими Бернулиеви опита с вероятност за успех p. Ако с k означим брой успехи, то:

$$P(X = k) = \binom{n}{k} p^k q^{n-1}$$

$$EX = p, VarX = pq$$

> **R:**
>
> - `rbinom(m, n, prob)` генерира m експеримента, всеки от който се състои от n опита, всеки опит с вероятност за успех prob.
> - `dbinom(x, n, prob)` дава **P(X = x)**,
>   `dbinom(c(1,2,3,4,5,6), n, prob)` ще ни даде вектор от стойностите: [P(X = 1), …, P(X = 6)]
>   За да нарисуваме графиката на теоретичното разпределение в интервала [0, a] правим:
>   `y = dbinom(0:a, n, prob)`
>   `plot(0:a, y, type = "s")`
> - `pbinom(q, n, prob, lower.tail = TRUE)` намира **p** в уравнението **p = P(X < q)**
> - `qbinom(p, n, prob, lower.tail = TRUE)` намира **q** в уравнението **p = P(X < q)**

Геометрично разпределение

Геометричното разпределение е дискретно разпределение на броя опити до първи успех. Ако означим с k броя опити, то:

$$P(X = k) = pq^{k-1}$$

$$EX = 1/p, VarX = q/p^2$$

> **R:**
>
> - `rgeom(n, prob)` генерира за n експеримента за всеки от тях колко опита са били нужни да постигнем успех.
> - `dgeom(x, prob, log = FALSE)` дава **P(X = x)**
>   `dgeom(c(1,2,3,4,5,6), prob, log = FALSE)` ще ни даде вектор от стойностите: [P(X = 1), …, P(X = 6)]
> - `pgeom(q, prob, lower.tail = TRUE)` намира **p** в уравнението **p = P(X < q)**
> - `qgeom(p, prob, lower.tail = TRUE)` намира **q** в уравнението **p = P(X < q)**

Отрицателно биномно разпределение

Отрицателното биномно разпределение е дискретно разпределение на броя опити до r-ти успех. Ако с k означим броя опити, то:

$$P(X = k) = \binom{k-1}{r-1} p^r q^{k-r}$$

$$EX = r/p, VarX = rq/p^2$$

> **R:**
>
> - `rnbinom(n, r, prob)` генерира n опита от отрицателно биномно разпределение с параметрите r и prob
> - `dnbinom(x, r, prob)` дава **P(X = x)**
> - `pnbinom(q, r, prob, lower.tail = TRUE)` намира **p** в уравнението **p = P(X < q)**
> - `qnbinom(p, r, prob, lower.tail = TRUE)` намира **q** в уравнението **p = P(X < q)**

Поасоново разпределение

Поасоновото разпределение е дискретно и описва брой случили се независими събития и ние знаем средния брой случили се събития за интервал от време - това е

$$\lambda$$

Ако k е броят случили се събития, то:

$$P(X = k) = \lambda^k \epsilon^{-\lambda} / k!$$

$$EX = \lambda, VarX = \lambda$$

> **R:**
>
> - `rpois(n, lambda)` генерира за n експеримента средно във всеки експеримент колко пъти се е случило събитието.
> - `dpois(x, lambda, log = FALSE)` дава **P(X = x)**
> - `ppois(q, lambda, lower.tail = TRUE)` намира **p** в уравнението **p = P(X < q)**
> - `qpois(p, lambda, lower.tail = TRUE)` намира **q** в уравнението **p = P(X < q)**

Хипергеометрично разпределение

Ако имаме популация с N елемента, k от които имат дадено качество. Ако изберем n елемента от тази популация, P(X = x) ни дава каква е вероятността x от тези n елемента да имат качеството. Вероятността е брой благоприятни случаи, върху всички случаи и е:

$$P(X = k) = \binom{r}{k} \binom{N-r}{n-k} / \binom{N}{n}$$

$$EX = \frac{nr}{N}, VarX = \frac{nr}{N} \frac{N-r}{N} \frac{N-n}{N-1}$$

> **Забележка! В R имаме следното:** (от документацията)
>
> **x - number of white balls drawn without**
> **m - the number of white balls in the urn.**
> **n - the number of black balls in the urn.**
> **k - the number of balls drawn from the urn.**

Т.е. при нас N = m + n (всички топчета)
x := k
m := r

n := N - r
k := n

> **R:**
>
> - `rhyper(nn, m, n, k)` => В нашия запис имаме `rhyper(nn, r, N - r, n)`
>   където nn - number of observations. брой наблюдения (както обикновено първия аргумент на rxxx функция е колко на брой експерименти искаме да генерираме)
> - `dhyper(x, m, n, k)`
> - `phyper(q, m, n, k, lower.tail = TRUE)` намира **p** в уравнението **p = P(X < q)**
> - `qhyper(p, m, n, k, lower.tail = TRUE)` намира **q** в уравнението **p = P(X < q)**

Непрекъснати разпределения

Равномерно разпределение

В интервала [a, b] наблюдаваме константна вероятност.

$$f(x) = const, a <= x <= b$$

$$EX = (b + a)/2, VarX = (a - b)^2/12$$

> **R:**
>
> - `runif(n, min)`
> - `dunif(x, min, max)` P(X = x)
> - `punif(q, min, max, lower.tail = TRUE)` намира **p** в уравнението **p = P(X < q)**
> - `qunif(p, min, max, lower.tail = TRUE)` намира **q** в уравнението **p = P(X < q)**

Нормално разпределение

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}}\epsilon^{\dfrac{-(x-\mu)^2}{2\sigma^2}}$$

$$EX = \mu, VarX = \sigma^2$$

> **R:**
>
> - `rnorm(n, mean = 0, sd = 1)` Генерира n опита (пример: генерираме n оценки на ученици, като знаем средната оценка и дисперсията)
> - `dnorm(x, mean = 0, sd = 1)` P(X = x)
> - `pnorm(q, mean = 0, sd = 1, lower.tail = TRUE)` намира **p** в уравнението **p = P(X < q)**
> - `qnorm(p, mean = 0, sd = 1, lower.tail = TRUE)` намира **q** в уравнението **p = P(X < q)**

Как чертаем графиката в интервала [a, b] със средно mu и стандартно отклонение stand?
x = seq(from = a, to = b by = 0.1)
y = dnorm(x, mean = mu, sd = stand)
plot(x, y, type = “l”)

Експопенциално разпределение

Живот на електроуред.
Времето до поява на първо събитие също е експоненциално разпределено.

$$f(x) = \lambda\epsilon^{-\lambda x}, x >= 0$$

$$f(x) = 0, x < 0$$

$$EX = 1/\lambda, VarX = 1/\lambda^2$$

> **R:**
>
> - `rexp(n, rate)`
> - `dexp(x, rate)`
> - `pexp(q, rate, lower.tail = TRUE)`
> - `qexp(p, rate, lower.tail = TRUE)`
>
>   Ако средното е mu, то rate = 1/mu.
>   Пример 1 събитие на година, търсим вероятността да изминат поне 3 **месеца** до земетресение? => Ще гледаме месеци - средно се случва 1/12 земетресение в месеца => rate = 1/12, вероятността е P(T > 3) = e^(-lambda * t) = e ^(-1/12 * 3)