ТОМСКИЙ
ПОЛИТЕХНИЧЕСКИЙ
УНИВЕРСИТЕТ

Статистическое моделирование и прогнозирование
Лекция 1. Случайные величины и законы распределения

Семёнов Михаил Евгеньевич
к. ф.-м. н., доцент ОЭФ ИЯТШ

Томский политехнический университет
16 марта 2020 г.

Содержание

Случайные величины и законы распределения

Случайная величина X представляет собой однозначную действительную функцию, заданную на пространстве элементарных событий Ω. Каждая случайная величина задаёт распределение вероятностей на множестве своих возможных значений.

Законом распределения случайной величины X называется всякое соотношение, устанавливающее связь между возможными значениями этой случайной величины и соответствующими им вероятностями. Случайная величина X считается заданной, если известен её закон распределения.

Функция распределния вероятностей

Наиболее общей формой закона распределения является *функция распределения вероятностей* случайной величины, определяемая равенством

$$F(x) = \mathbb{P}\{x < X\}. \tag{1}$$

Основные свойства функции распределения $F(x)$

- *Значения* функции распределения ограничены интервалом: $0 \le F(x) \le 1$;
- Функция распределения — неубывающая функция: $F(x_2) \ge F(x_1)$, если $x_2 > x_1$;
- *Предельные значения* аргумента соответствуют предельным значениям функции распределения: $F(-\infty) = 0$, $F(\infty) = 1$;
- *Вероятность* события $X \in [\alpha, \beta)$ равна приращению функции распределения на интервале: $\mathbb{P}\{\alpha \le X < \beta\} = F(\beta) - F(\alpha)$.

Дискретные случайные величины

Дискретной называется случайная величина, множество возможных значений которой конечное или счётное. В качестве закона распределения дискретной случайной величины часто используют ряд распределения, записываемый в виде таблицы $2 \times n$:

$$X \sim \begin{pmatrix} x_1 & x_2 & \dots & x_n \\ & & & \\ p_1 & p_2 & \dots & p_n \end{pmatrix}, \tag{2}$$

где $p_i = \mathbb{P}\{X = x_i\}$ при этом $\sum_{i=1}^{n} p_i = 1$.
Функция распределения дискретной случайной величины будет иметь разрывы первого рода (скачки), в точках, соответствующих значениям случайной величины x_i (абсциссы скачков). Причем величины этих скачков будут равны вероятностям соответствующих значений p_i (ординаты скачков).

Непрерывные случайные величины

Непрерывной называется случайная величина, имеющая непрерывную и дифференцируемую функцию распределения $F(x)$. В качестве закона распределения непрерывной случайной величины обычно используется функция плотности распределения вероятностей:

$$f(x) = \frac{dF(x)}{dx}. \tag{3}$$

Основные *свойства* плотности распределения вероятностей $f(x)$

- Плотность распределения вероятностей — функция неотрицательная: $f(x) \geq 0$;
- Плотность распределения удовлетворяет *условию нормировки*: $\int_{-\infty}^{\infty} f(x)dx = 1$;
- *Вероятность* события $X \in [\alpha, \beta]$ равна интегралу на соответствующем отрезке от плотности распределения: $\mathbb{P}\{\alpha \leq X \leq \beta\} = \int_{\alpha}^{\beta} f(x)dx$;

Наиболее распространённые распределения

Законы распределения:

- биномиальный
- геометрический
- равномерный
- показательный
- нормальный
- логнормальный
- Пуассона
- Пирсона
- Стьюдента
- Фишера
- Лапласа
- Коши

Наиболее распространённые распределения

Дискретная случайная величина X имеет *биномиальное* распределение с параметрами $n \in \mathbb{Z}^+$, $p : X \sim B(n, p)$, если она принимает целочисленные значения $k = 0, 1, \ldots, n$ с вероятностями, определяемыми формулой Бернулли

$$p_k = \mathbb{P}\{X = k\} = C_n^k p^k q^{n-k}, \tag{4}$$

где $C_n^k = \frac{n!}{k!(n-k)!}$, $p \in (0, 1)$, $p = 1 - q$.

Математическое ожидание и дисперсия случайной величины, подчиняющейся биномиальному закону $X \sim B(n, p)$ вычисляются по формулам:
$M(X) = np$, $D(X) = npq$.

Наиболее распространённые распределения

Нормальное распределение обычно возникает при рассмотрении суммы большого количества независимо распределённых случайных величин с конечной дисперсией. Непрерывная случайная величина X имеет нормальное распределение: $X \sim N(a, \sigma)$ если её плотность вероятности имеет вид:

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left\{-\frac{(x-a)^2}{2\sigma^2}\right\} = \varphi\left(\frac{x-a}{\sigma}\right), \tag{5}$$

где $x, a \in \mathbb{R}$, $\sigma > 0$, $\varphi(z)$ – функция Гаусса, определяемая равенством $\varphi(z) = \frac{1}{\sqrt{2\pi}} \exp(\frac{-z^2}{2})$.

Математическое ожидание и дисперсия случайной величины, подчиняющейся нормальному закону $X \sim N(a, \sigma)$ вычисляются по формулам: $M(X) = a$, $D(X) = \sigma^2$.

«Экзотические» распределения

Для того, чтобы отразить такие свойства финансовых временных рядов, как *тяжёлые хвосты* или *асимметрия* могут быть использованы следующие распределения:

- гиперболическое;
- устойчивое;
- Мейкснера.

Гиперболическое распределение

Функция плотности гиперболического распределения [Barndorff-Nielsen and Blæsild, 1983]:

$$f_H(x|\pi,\zeta,\delta,\mu) = \frac{1}{2\sqrt{1+\pi^2}K_1(\zeta)} e^{-\zeta\left[\sqrt{1+\pi^2}\sqrt{1+\left(\frac{x-\mu}{\delta}\right)^2}-\pi\frac{x-\mu}{\delta}\right]}, \tag{6}$$

где

- $K_1(x)$ — модифицированная функция Бесселя третьего рода 1-го порядка [Bessel, 1824],
- $\pi \in \mathbb{R}$, $\zeta > 0$, $\delta > 0$, $\mu \in \mathbb{R}$.

Устойчивое распределение

Характеристическая функция устойчивого распределения [Nolan, 2009]:

$$\varphi_S(x|\alpha,\beta,\gamma,\delta) = \exp\left[ix\delta - |\gamma x|^{\alpha}\left(1 - i\beta \operatorname{sgn}(x)\Phi(x)\right)\right], \quad (7)$$

$$\Phi(x) = \begin{cases} \left(|\gamma x|^{1-\alpha} - 1\right)\tan\dfrac{\pi\alpha}{2}, & \alpha \neq 1, \\ -\dfrac{2}{\pi}\log|\gamma x|, & \alpha = 1, \end{cases}$$

где

- $\alpha \in (0;\ 2]$, $\beta \in [-1;\ 1]$, $\gamma > 0$, $\delta \in \mathbb{R}$,
- i — мнимая единица.

Распределение Мейкснера

Функция плотности распределения Мейкснера [Schoutens, 2002]:

$$f_M(x|\alpha,\beta,\delta,\mu) = \frac{\left(2\cos\frac{\beta}{2}\right)^{2\delta}}{2\alpha\pi\Gamma(2\delta)} \exp\frac{\beta(x-\mu)}{\alpha} \left|\Gamma\left(\delta + i\frac{x-\mu}{\alpha}\right)\right|^2, \tag{8}$$

где

- $\Gamma(z)$ — гамма-функция с комплексным аргументом z,
- $\alpha > 0$, $|\beta| < \pi$, $\delta > 0$, $\mu \in \mathbb{R}$.

Постановка заданий

- Построить графики плотности вероятностей и функции распределения $F(x)$ случайной величины X распределённой а) дискретно, б) непрерывно. Значения параметров распределения (биномиальный, геометрический, равномерный, показательный, нормальный, логнормальный, Пуассона, Пирсона, Стьюдента, Фишера) указать самостоятельно. Сохранить скрипт в файле (*.r или *.py), одно распределение – один файл.
- Подписать оси графиков, вывести два графика в одном графическом окне.
- Подготовить инструкции использования функций: par(), layout(), windows(), plot().
- Выборку, подчиняющуюся экспоненциальному распределению ($\lambda = 3$), трансформировать в выборку, подчиняющуюся стандартному нормальному распределению $N(0, 1)$. Провести тест на нормальность.
- Построить функцию распределения дискретной случайной величины из парадокса дня рождений.

Результат, R

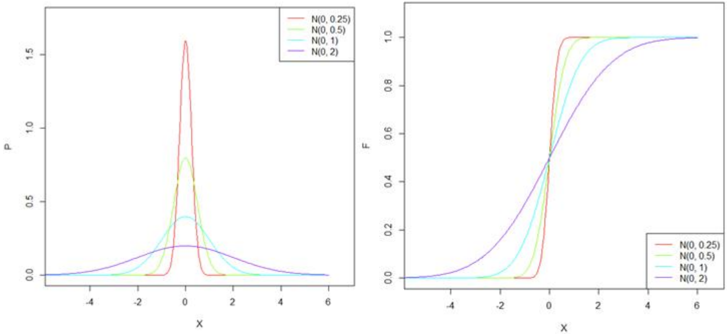

Рисунок 1 – Графики плотности вероятностей (слева) и функции распределения (справа), $X \sim N(a, \sigma)$

Результат, Python

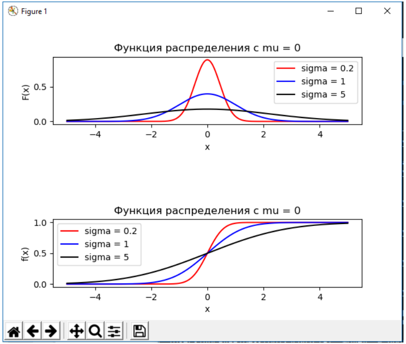

Рисунок 2 – Графики функции распределения (вверху) и плотности вероятностей (внизу), $X \sim N(a, \sigma)$

Скрипт на R [Буховец et al., 2010]

Алгоритм 1 Графики плотности вероятностей и функции распределения случайной величины

Вход: вероятность успеха p, число испытаний n.

```
source("probGraph.r")
p <- seq(1, 7, 2)/10
n <- 12; x <- seq(0, n)
P <- sapply(p, function(pp) dbinom(x, n, pp))
F <- sapply(p, function(pp) pbinom(x, n, pp))
lab <- sapply(p, function(pp) sprintf("B(%.0f, %.3g) n, pp))
dgraph(x, P, lab)
pgraph(x, F, lab)
```

Выход: Графики.

Скрипт на Phyton, Arsalan Dorzhiev

Алгоритм 2 Графики

Вход: математическое ожидание, дисперсия.

```
import math, scipy.stats as st, numpy as np, matplotlib.pyplot as plt
plt.subplot (3, 1, 1)
x = np.linspace(-5, 5,1000)
p1, = plt.plot(x, st.norm.pdf(x,0, math.sqrt(0.2)), color = ’red’)
p2, = plt.plot(x, st.norm.pdf(x,0, math.sqrt(1)), color = ’blue’)
p3, = plt.plot(x, st.norm.pdf(x,0, math.sqrt(5)), color = ’black’)
plt.legend([p1, p2, p3], ["sigma = 0.2 "sigma = 1 "sigma = 5"])
plt.xlabel(’x’) # обозначение оси абсцисс
plt.ylabel(’F(x)’)
plt.title(’Функция распределения с mu = 0’)
plt.subplot (3, 1, 3)
p1, = plt.plot(x, st.norm.cdf(x,0, math.sqrt(0.2)), color = ’red’)
p2, = plt.plot(x, st.norm.cdf(x,0, math.sqrt(1)), color = ’blue’)
p3, = plt.plot(x, st.norm.cdf(x,0, math.sqrt(5)), color = ’black’)
plt.legend([p1, p2, p3], ["sigma = 0.2 "sigma = 1 "sigma = 5"])
plt.xlabel(’x’)   обозначение оси абсцисс
plt.ylabel(’f(x)’)
```

Преобразование распределения случайной величины

Алгоритм 3 Трасформация распределения случайной величины

Вход: Случайная величина, распределенная экспоненциально

```
set.seed(1234) # for reproducibility
n = 1000; lam = 3
x = rexp(n, lam)
mean(x); sd(x)
u = 1 - exp(-lam*x)
mean(u); var(u); # 1/12
z = qnorm(u)
mean(z); sd(z)
par(mfrow=c(1,3))
hist(x); hist(u); hist(z);
shapiro.test(z)
```

Выход: Случайная величина, распределенная нормально

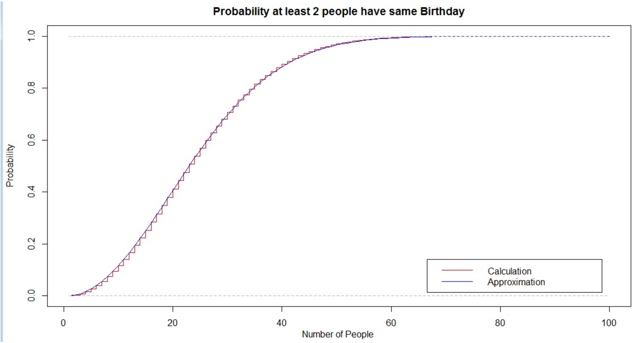

Рисунок 3 – Функции распределения для парадокса дней рождения

Barndorff-Nielsen, O. and Blæsild, P. (1983).
Hyperbolic distributions.
Encyclopedia of Statistical Sciences, 3:700–707.

Bessel, F. (1824).
Untersuchung des Theils der planetarischen Störungen.
Berlin Abhandlungen.

Nolan, J. P. (2009).
Stable Distributions – Models for Heavy Tailed Data.
Birkhauser, Boston.

Schoutens, W. (2002).
Meixner processes: Theory and applications in finance.
Eurandom Report 2002-004, Eindhoven, Eindhoven, Netherlands.

Буховец, , Москалев, , Богатова, , and Бирючинская, (2010).
Статистический анализ данных в системе R. Учебное пособие.
ВГАУ, Воронеж.