УДК 519.24

ОЦЕНКА ЧИСЛОВЫХ ХАРАКТЕРИСТИК И ВИЗУАЛИЗАЦИЯ РАСПРЕДЕЛЕНИЯ ПУАССОНА (С ИСПОЛЬЗОВАНИЕМ АЛГОРИТМИЧЕСКОГО ЯЗЫКА R)

Борис Игнатьевич Смагин

доктор экономических наук, профессор

bismagin@mail.ru

Мичуринский государственный аграрный университет

г. Мичуринск, Россия

Аннотация. Распределение Пуассона является одним из наиболее распространенных, в первую очередь, при моделировании систем массового обслуживания. В статье изложена логика формирования данного закона, определены его числовые характеристики (математическое ожидание и дисперсия). Использование алгоритмического языка R позволяет автоматизировать вычисления и визуализировать как плотность, так и функцию распределения данной случайной величины.

Ключевые слова: испытания Бернулли, распределение Пуассона, математическое ожидание, дисперсия, алгоритмический язык R.

Во многих приложениях имеются испытания Бернулли, в которых число испытаний n относительно велико, а вероятность «успеха» p очень мала. При этом их произведение np = λ имеет конечное значение. В таких случаях для биномиального распределения удобно использовать предложенное Пуассоном приближение.

$$P(k,\lambda)=C_n^k\cdot p^k\cdot q^{n-k}=\frac{n(n-1)\cdot(n-2)\cdot\ldots\cdot(n-k+1)}{k!}\cdot p^k\cdot(1-p)^{n-k}.$$

Так как np = λ, то p = λ/n. Следовательно

$$P(k,\lambda)=\frac{n(n-1)\cdot(n-2)\cdot\ldots\cdot(n-k+1)}{k!}\cdot\left(\frac{\lambda}{n}\right)^k\cdot\left(1-\frac{\lambda}{n}\right)^{n-k}.$$

Приняв во внимание, что n имеет очень большое значение, найдем предел

$$\lim_{n\to\infty}P(k,\lambda)=\lim_{n\to\infty}\frac{n(n-1)\cdot(n-2)\cdot\ldots\cdot(n-k+1)}{k!}\cdot\frac{\lambda^k}{n^k}\cdot\left(1-\frac{\lambda}{n}\right)^{n-k}=$$

$$=\frac{\lambda^k}{k!}\cdot\lim_{n\to\infty}\left[1\left(1-\frac{1}{n}\right)\left(1-\frac{2}{n}\right)\cdot\ldots\cdot\left(1-\frac{k-1}{n}\right)\left(1-\frac{\lambda}{n}\right)^{n-k}\right]=$$

$$=\frac{\lambda^k}{k!}\cdot\lim_{n\to\infty}\left(1-\frac{\lambda}{n}\right)^n\cdot\lim_{n\to\infty}\left(1-\frac{\lambda}{n}\right)^{-k}=\frac{\lambda^k}{k!}\cdot e^{-\lambda}\cdot1=\frac{\lambda^k}{k!}\cdot e^{-\lambda}.$$

Это и есть классическое пуассоновское приближение для биномиального распределения.

Рассмотрим дискретную случайную величину X, которая может принимать только целые, неотрицательные значения: 0, 1, 2,…., k,…, причем последовательность этих значений теоретически не ограничена. Говорят, что случайная величина X распределена по закону Пуассона, если вероятность того, что она примет определенное значение k, выражается формулой

$$P_k=\frac{\lambda^k}{k!}e^{-a}\quad(k=0,1,\ldots),$$

где λ – некоторая положительная величина, называемая параметром закона Пуассона. Убедимся прежде всего, что последовательность вероятностей, задаваемая данной формулой, может представлять собой ряд распределения, т.е. что сумма всех вероятностей P_k равна единице. Имеем:

$$\sum_{k=0}^{\infty} P_k = \sum_{k=0}^{\infty} \frac{\lambda^k}{k!} e^{-\lambda} = e^{-\lambda} \sum_{k=0}^{\infty} \frac{\lambda^k}{k!} = 1$$

Определим основные характеристики – математическое ожидание и дисперсию случайной величины X, распределенной по закону Пуассона. По определению математического ожидания

$$M\left\{X\right\} = \sum_{k=0}^{\infty} kP_k = \sum_{k=0}^{\infty} k \frac{\lambda^k}{k!} e^{-\lambda}$$

Первый член суммы (соответствующий k = 0) равен нулю, следовательно, суммирование можно начинать с k = 1:

$$M\left\{X\right\} = \sum_{k=1}^{\infty} k \frac{\lambda^k}{k!} e^{-\lambda} = \lambda e^{-\lambda} \sum_{k=1}^{\infty} \frac{\lambda^{k-1}}{(k-1)!}$$

Обозначим k – 1 = m; тогда

$$M\left\{X\right\} = \lambda e^{-\lambda} \sum_{m=0}^{\infty} \frac{\lambda^m}{m!} = \lambda e^{-\lambda} e^{\lambda} = \lambda.$$

Таким образом, параметр λ представляет собой не что иное, как математическое ожидание случайной величины X.

Для определения дисперсии будем использовать формулу $D\{X\} = M\{X^2\} - [M\{X\}]^2$. Найдем $M\{X^2\}$:

$$M\left\{X^2\right\} = \sum_{k=0}^{\infty} k^2 \frac{\lambda^k}{k!} e^{-\lambda} = \lambda \sum_{k=1}^{\infty} k \frac{\lambda^{k-1}}{(k-1)!} e^{-\lambda} =$$

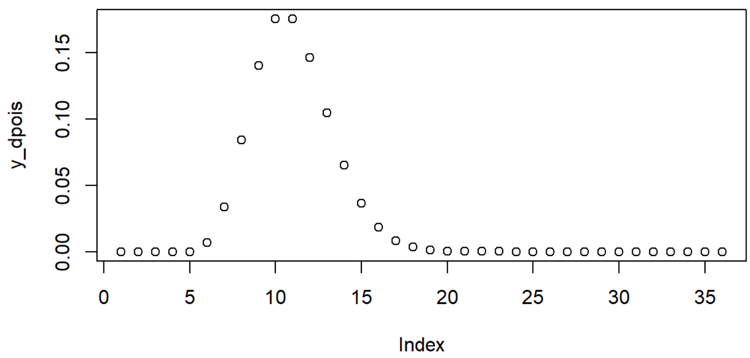
$$= \lambda \sum_{k=1}^{\infty} \left[(k-1)+1\right] \frac{\lambda^{k-1}}{(k-1)!} e^{-\lambda} =$$

$$= \lambda \left[\sum_{k=1}^{\infty} (k-1) \frac{\lambda^{k-1}}{(k-1)!} e^{-\lambda} + \sum_{k=1}^{\infty} \frac{\lambda^{k-1}}{(k-1)!} e^{-\lambda} \right] = \lambda(\lambda+1).$$

Тогда $D\{X\} = M\{X^2\} - [M\{X\}]^2 = \lambda(\lambda+1) - \lambda = \lambda$.

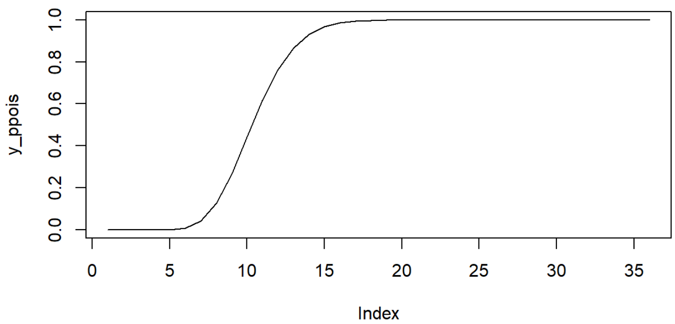
Чтобы создать плотность Пуассона в R, нам сначала нужно создать последовательность из целых чисел. Затем мы возвращаем соответствующие значения плотности Пуассона для каждого из этих значений [1]. В примере мы используем лямбда-выражение, равное 5. Если мы хотим создать график, показывающий эти значения плотности вероятности, мы применяем функцию plot.

```
plot(y_dpois)
x_dpois <-seq(-5,30,by=1)
> y_dpois <-dpois(x_dpois,lambda = 5)
> plot(y_dpois)
```



Аналогичная процедура используется для построения кумулятивной функции распределения (CDF) распределения Пуассона.

```
x_ppois <-seq(-5,30,by=1)
> y_ppois <-ppois(x_ppois,lambda = 5)
> plot(y_ppois,type = 'l')
```



Следует отметить, что распределение Пуассона находит широкое применение в теории массового обслуживания, основным элементом которой является входящий поток требований.

Входящим потоком требований мы будем называть число требований, поступивших за промежуток времени (0;t); это неубывающий случайный процесс X(t), принимающий только целочисленные значений (0,1,2,… и т.д).

Случайный поток называется стационарным, если для любой группы из конечного числа непересекающихся отрезков времени длительностью t_1, t_2, …, t_n вероятность появления в них соответственно k_1, k_2, …, k_n требований зависит только от этих чисел и от длин указанных промежутков, но не от их расположения на оси времени.

Отсутствие последействия означает, что вероятность поступления за отрезок времени длительности t определенного числа требований не зависит от того, сколько требований уже поступило в систему, т.е. не зависит от предыстории изучаемого явления. Случайный процесс без последействия также называют марковским процессом.

Ординарность потока требований выражает собой условие практической невозможности появления более одного требования в один и тот же момент времени.

Входящий поток, удовлетворяющий требованиям стационарности, отсутствия последействия и ординарности называют простейшим потоком.

Доказана следующая теорема [2].

Теорема. Для простейшего потока

$$P_k(t) = \frac{(\lambda t)^k}{k!} \cdot e^{-\lambda t}, \ k = 0,1,2,...$$

Таким образом, входящий поток, удовлетворяющий требованиям стационарности, отсутствия последействия и ординарности подчиняется распределению Пуассона.

Список литературы:

1. Зарядов И.С. Статистический пакет R: Теория вероятностей и математическая статистика // М.: РУДН. 2010. 141с.

2. Смагин Б.И. Экономико-математические методы: учебник для академического бакалавриата // М.: Издательство Юрайт. 2017. 272с.

UDC 519.24

ESTIMATION OF NUMERICAL CHARACTERISTICS AND VISUALIZATION OF THE POISSON DISTRIBUTION (USING THE ALGORITHMIC LANGUAGE R)

Boris I. Smagin
doctor of economics, professor
bismagin@mail.ru
Michurinsk State Agrarian University
Michurinsk, Russia

Abstract. Poisson distribution is one of the most common, first of all, when modeling queuing systems. The article describes the logic of the formation of this law, defines its numerical characteristics (mathematical expectation and variance). Using the algorithmic language R allows you to automate calculations and visualize both the density and the distribution function of a given random variable.

Keywords: Bernoulli tests, Poisson distribution, mathematical expectation, variance, algorithmic language

Статья поступила в редакцию 01.02.2024; одобрена после рецензирования 20.03.2024; принята к публикации 22.03.2024.

The article was submitted 01.02.2024; approved after reviewing 20.03.2024; accepted for publication 22.03.2024.