ЛАБОРАТОРНЫЙ ПРАКТИКУМ ПО МАТЕМАТИЧЕСКИМ ОСНОВАМ КИБЕРНЕТИКИ

Лабораторная работа

«Изучение распределений дискретных случайных величин в среде MATHCAD»

Томск 2009

Лабораторный практикум предназначен для выполнения студентами направления 220200 «Автоматизация и управление» и студентами специальности 220301 «Автоматизация технологических процессов и производств (в нефтегазовой отрасли)» цикла лабораторных работ по дисциплине «Математические основы кибернетики» / сост., В.А. Рудницкий, Томск, ТПУ, каф. ИКСУ АВТФ, 2009 г.

Целью лабораторной работы является изучение распределений дискретных случайных величин с помощью математического пакета Mathcad.

Теория вероятностей изучает математические модели случайных явлений окружающего нас мира. Одно из центральных понятий теории вероятностей – понятие случайной величины. *Случайной величиной* называется числовая функция, заданная на множестве случайных событий.

Например, случайной величиной является число очков, выпавших при бросании игральной кости, или рост случайно выбранного из учебной группы студента. В первом случае мы имеем дело с *дискретной* случайной величиной (она принимает значения из дискретного числового множества, т.е. множества, состоящего из конечного или счетного числа элементов); во втором случае – с *непрерывной* случайной величиной (она принимает значения из непрерывного числового ряда).

Каждая случайная величина полностью определяется своей функцией распределения.

Если ξ - случайная величина, то функция $\mathbf{F(x)= F_{\xi}\ (x)= P(\xi<x)}$ называется *функцией распределения* случайной величины ξ. Здесь $\mathbf{P(\xi<x)}$ – вероятность того, что случайная величина ξ принимает значение меньшее x.

Функция распределения любой случайной величины обладает следующими свойствами:

- **F(x)** определена на всей числовой прямой;
- **F(x)** не убывает, т.е. если $\mathbf{x_1 \le x_2}$, то $\mathbf{F(x_1) \le F(x_2)}$;
- $\mathbf{F(-\infty)=0}$ **и** $\mathbf{F(\infty)=1}$, т.е. $\lim\limits_{x\to-\infty} \mathbf{F(x)} = 0$ и $\lim\limits_{x\to\infty} \mathbf{F(x)} = 1$;
- F(x) непрерывна слева , т.е $\lim\limits_{x\to x_0-0} \mathbf{F(x)} = \mathbf{F(x_0)}$.

Важно понимать, что функция распределения является «паспортом» случайной величины, она содержит всю информацию об этой случайной величине, и потому изучение случайной величины заключается в исследовании ее *функции распределения,* которую часто называют просто *распределением*. Например, когда говорят о *нормальном распределении*, то подразумевают случайную величину, имеющую *нормальную функцию распределения*.

Вероятность того, что значение случайной величины ξ попадает в интервал **(a, b)** для дискретной случайной величины вычисляется по формуле.

$$\mathbf{P(a<\xi<b)} = \sum_{x_i \in (a,b)} \mathbf{p_i}.$$

Прежде чем приступать к решению задач теории вероятностей с помощью математического пакета Mathcad, необходимо ознакомиться с инструментами, которые он предоставляет для их решения.

У дискретной случайной величины функция распределения ступенчатая Известно, что дискретная случайная величина ξ, принимающая значения $x_1<x_2<\ldots<x_i<\ldots$ с вероятностями $p_1, p_2, \ldots, p_i, \ldots,$ может быть задана распределением – таблицей вида

ξ	x_1	x_2	…	x_i	…	x_n
$\boldsymbol{p}$	p_1	p_2	…	p_i	…	p_n

Такая таблица называется *рядом распределения дискретной случайной величины* или просто *распределением случайной величины*. В среде Mathcad их удобно хранить в виде матриц размерности 2×n. В практических задачах именно такая форма представления распределения наиболее удобна.

Функция распределения случайной величины, имеющей приведенное выше распределение, имеет вид

$$\mathbf{F(x)} = \begin{cases} \mathbf{0}, & \mathbf{x < x_1}, \\ \mathbf{p_1}, & \mathbf{x_1 \le x < x_2}, \\ \mathbf{p_1 + p_2}, & \mathbf{x_2 \le x < x_3}, \\ \dots\dots\dots\dots\dots\dots & \\ \mathbf{p_1 + p_2 + \dots + p_{n-1}}, & \mathbf{x_{n-1} \le x < x_n} \\ \mathbf{1}, & \mathbf{x_n \le x} \end{cases}.$$

Рассмотрим фрагмент рабочего документа Mathcad с определением распределения случайной величины, ее функции распределения и графиком функции распределения для случайной величины, имеющей следующее распределение:

ξ	1	0	7	4	-2
$\boldsymbol{p}$	0.1	0.5	0.1	0.1	0.2

$$\text{ORIGIN} := 1 \qquad A := \begin{pmatrix} -2 & 0 & 1 & 4 & 7 \\ 0.2 & 0.5 & 0.1 & 0.1 & 0.1 \end{pmatrix}$$

Функция распределения случайной величины может быть задана двумя основными способами:

$$F(x) := \begin{array}{|l} 0 \ \text{if} \ -\infty < x < A_{1,1} \\ A_{2,1} \ \text{if} \ A_{1,1} \le x < A_{1,2} \\ \left(\left(A_{2,1} + A_{2,2}\right)\right) \ \text{if} \ A_{1,2} \le x < A_{1,3} \\ \left(A_{2,1} + A_{2,2} + A_{2,3}\right) \ \text{if} \ A_{1,3} \le x < A_{1,4} \\ \left(A_{2,1} + A_{2,2} + A_{2,3} + A_{2,4}\right) \ \text{if} \ A_{1,4} \le x < A_{1,5} \\ 1 \ \text{if} \ A_{1,5} \le x < \infty \end{array}$$

$$G(x) := \begin{array}{|l} 0 \ \text{if} \ -\infty < x < -2 \\ 0.2 \ \text{if} \ -2 \le x < 0 \\ 0.2 + 0.5 \ \text{if} \ 0 \le x < 1 \\ 0.2 + 0.5 + 0.1 \ \text{if} \ 1 \le x < 4 \\ 0.2 + 0.5 + 0.1 + 0.1 \ \text{if} \ 4 \le x < 7 \\ 1 \ \text{if} \ 7 \le x < \infty \end{array}$$

Графики функции распределения случайной величины представлены на Рисунке 1:

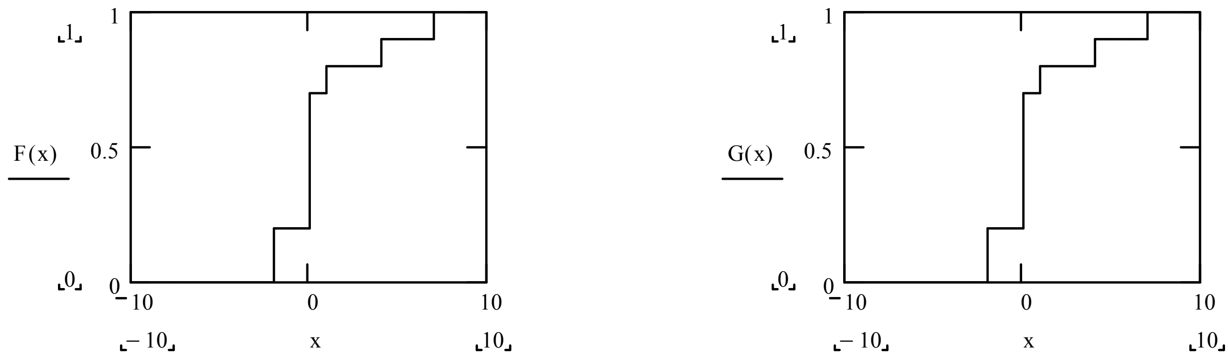

Рисунок 1.- Графики функции распределения случайной величины.

В рабочем документе распределение случайной величины задано в матрице **A**, где $\mathbf{A_{1,i}}$ - значения случайной величины, $\mathbf{A_{2,i}}$ - соответствующий вероятности, i=**1, 2, 3, 4, 5**. Функцию распределения, заданную различными выражениями на разных интервалах изменения аргумента, можно определить следующим образом:

1. Развернуть панель программных элементов щелчком по кнопке «Programming Toolbar» и панель знаков отношений щелчком по кнопке «Boolean Toolbar». Не закрывать эти панели до окончания определения функции!
2. Ввести имя функции переменной x и знак присваивания (панель «Calculator Toolbar» → «Definition»).
3. В панели «Programming Toolbar» щелкнуть по кнопке «Add Line», ввести в помеченной позиции 0; щелкнуть по кнопке «if» и ввести неравенства, определяющие первый интервал изменения аргумента x. Символ ∞ можно ввести щелчком по соответствующей кнопке в панели «Calculus Toolbar».
4. Аналогичные пункту 3 операции повторить для остальных интервалов изменения аргумента **x** функции **F(x).**

В рабочем документе приведены два способа определения функции распределения случайной величины: с использованием имен переменных и с использованием конкретных значений этих переменных. Графики функций построены стандартным для декартовых графиков способом «Graph Toolbar»→ «X-Y Plot». «Programming Toolbar». Следует учитывать то обстоятельство, что Mathcad не совсем корректно строит графики ступенчатых функций, соединяя отрезками значения функции в точке скачка. Более точный график функции распределения представляет собой отрезки, параллельные оси абсцисс с «выколотым» правым концом.

Для операций со случайными величинами (непрерывными и дискретными) в Mathcad имеется библиотека встроенных функций наиболее распространенных стандартных распределений. Каждое распределение представлено тремя функциями – плотностью вероятностей, функцией распределения и функцией, обратной к функции распределения.

Например, для работы с нормальным распределением предназначены функции dnorm(x,μ,σ), pnorm(x,μ,σ) и qnorm(p,μ,σ). Значением функции dnorm(x,μ,σ) является значение в точке **x** плотности вероятностей случайной величины **ξ**, имеющей нормальное распределение с математическим ожиданием **Mξ=μ** и дисперсией $\mathbf{D\xi=\sigma^2}$. Значение функции pnorm(x,μ,σ) – значение функции распределения этой же случайной величины **ξ**. Значением функции qnorm(x,μ,σ) служит решение уравнения **F(x)=p**, где **F(x)** – функция распределения, определенная функцией pnorm(x,μ,σ), т.е. значением pnorm(x,μ,σ) является квантиль уровня p нормально распределенной случайной величины.

Можно заметить, что имена всех встроенных функций Mathcad, определяющих плотности вероятностей, начинаются с буквы **d**, определяющих функции распределения – с буквы **p**, определяющих квантили – с буквы **q**.

Ниже приведен список всех распределений представленных в библиотеке Mathcad:

1. Бета-распределение – dbeta(x,s_1,s_2), pbeta(x,s_1,s_2), qbeta(p,s_1,s_2).
2. Биномиальное распределение – dbinom(k,n,p), pbinom(k,n,p), qbinom(p,n,r).
3. Распределение Коши – dcauchy(x,l,s), pcauchy(x,l,s), qcauchy(p,l,s).
4. χ^2 –распределение – dchisq(x,d), pchisq(x,d), qchisq(p,d).
5. Экспоненциальное распределение – dexp(x,r), pexp(x,r), qexp(p,r).
6. Распределение Фишера (F-распределение)- dF(x,d_1,d_2), pF(x,d_1,d_2), qF(p,d_1,d_2)
7. Гамма-распределение – dgamma(x,s), pgamma(x,s), qgamma(p,s).
8. Геометрическое распределение – dgeom(x,p), pgeom(x,p), qgeom(p,r).
9. Логнормальное распределение - dlnorm(x,μ,σ), plnorm(x,μ,σ), qlnorm(p,μ,σ).
10. Логистическое распределение – dlogis(x,l,s), plogis(x,l,s), qlogis(p,l,s).
11. Отрицательное биномиальное распределение- dnbinom(k,n,p), pnbinom(k,n,p), qnbinom(p,n,r).
12. Нормальное распределение - dnorm(x,μ,σ), pnorm(x,μ,σ) , qnorm(p,μ,σ).
13. Распределение Пуассона – dpois(x, λ), ppois(x, λ), qpois(p, λ).
14. Распределение Стьюдента – dt(x,d), pt(x,d), qt(p,d).
15. Равномерное распределение – dunif(x,a,b), punif(x,a,b), qunif(p,a,b).
16. Распределение Вейбулла – dweibull(x,s), pweibull(x,s), qweibull(p,s).

С описанием и свойствами этих распределений можно ознакомится в разделе «Mathcad Help».

Кроме того, для вычисления числовых характеристик дискретных и непрерывных случайных величин применяются операторы интегрирования и дифференцирования, вычисления конечных сумм и суммирования рядов, которые вызываются по кнопке в панели «Calculus Toolbar» и заполнением соответствующих полей.

Рассмотрим более подробно наиболее распространенные распределения дискретных случайных величин, используемых при решении практических задач.

1. Биномиальное распределение (схема Бернулли).

Пусть проводится серия из n независимых испытаний, каждое из которых заканчивается либо «успехом», либо «неуспехом». Независимым называют испытания, исход каждого из которых не завистит от исхода остальных испытаний. Пусть в каждом испытании вероятность успеха **p**, а вероятность неудачи **q=1-p**. С таким испытанием можно связать случайную величину ξ, равную числу успешных опытов в серии из **n**-испытаний. Эта величина принимает целые значения от **0** до **n**. Ее распределение называется *биномиальным* и определяется формулой Бернулли

$$p_k = P(\xi = k) = C_n^k \cdot p^k \cdot q^{n-k},$$

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

Нетрудно убедиться, что $\sum\limits_{k=0}^{n} p_k = 1$.

В Mathcad для вычисления плотности вероятности и функции распределения случайной величины, имеющей биномиальное распределение, предназначены функции dbinom(k,n,p) и pbinom(k,n,p), значения которых – соответственно $\mathbf{p_k}$ и **F(k).**

2. Геометрическое распределение

Со схемой испытаний Бернулли можно связать еще одну случайную величину: η - число испытаний до первого успеха. Эта величина принимает бесконечное множество значений от **0** до ∞, и ее распределение определяется формулой

$$\mathbf{p_k=P(\eta=k)=q^k p,\quad k=0,1,\ldots,\ \ 0<p<1,\ q=1-p.}$$

используя формулу суммы бесконечно убывающей геометрической прогрессии, можно показать, что $\sum\limits_{k=0}^{\infty} p_k = 1$.

В Mathcad для вычисления плотности вероятности и функции распределения случайной величины, имеющей геометрическое распределение, предназначены функции dgeom(k,p) и pgeom(k,p), значения которых – соответственно $\mathbf{p_k}$ и **F(k).**

3.Гипергеометрическое распределение

Пусть в партии из **N** изделий имеется **M (M<N)** доброкачественных и **N-M** дефектных изделий. Если случайным образом из всей партии выбрать контрольную партию из n изделий, то число доброкачественных изделий в этой партии будет случайной величиной, которую обозначим ξ. Ее распределение имеет вид:

$$p_k = P(\xi = k) = \frac{C_M^k \cdot C_{N-M}^{n-k}}{C_N^n}, \qquad k = 0, 1, \ldots, \min(n, M),$$

и называется гипергеометрическим. Для любых значений параметров , входящих в распределение $\sum\limits_{k=0}^{\min(n,M)} p_k = 1$.

В Mathcad для вычисления плотности вероятности и функции распределения случайной величины, имеющей гипергеометрическое распределение, предназначены функции dhypergeom(k,M,N,n) и phypergeom(k,M,N,n), значения которых – соответственно $\mathbf{p_k}$ и **F(k).**

4.Пуассоновское распределение

Пуассоновское распределение имеет случайная величина μ, принимающая значения **k=0, 1, 2, …** с вероятностями

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

где $\lambda>0$ – параметр пуассоновского распределения. При любых $\boldsymbol{\lambda>0}$ $\sum\limits_{k=0}^{\infty} p_k = 1$.

В Mathcad для вычисления плотности вероятности и функции распределения случайной величины, имеющей пуассоновское распределение, предназначены функции dpois(k, λ) и ppois(k, λ), значения которых – соответственно $\mathbf{p_k}$ и **F(k).**

ЗАДАНИЕ

Для указанных значений параметров (Таблица 1.) вычислить и построить биномиальное распределение для серии из **n** независимых испытаний с вероятностью успеха **p**, пуассоновское распределение с параметром λ, геометрическое распределение с параметрами **n, p**, гипергеометрическое распределение с параметрами **N, M, n**.
Для каждого распределения выполнить следующее:

- Проверить равенство $\sum\limits_{k} p_k = 1$, где $p_k = P(\xi = k)$;
- Найти значение **k** для которого величина $P(\xi = k)$ максимальная. Исследовать зависимость этой вероятности от параметров распределения;
- Построить графики распределения и функции распределения;

График распределения- ломаная линия, вершинами которой являются точки (k, p_k)*, где k- значение случайной величины, а* p_k *– вероятность этого значения.*

- Вычислить вероятность попадания значений случайной величины интервал **(a, b).**

Для гипергеометрического распределения, с целью более детального изучения вопросов его практического использования, постройте в одних осях графики плотности распределения вероятности для двух случаев:

1) наличия большого числа бракованных изделий в партии деталей;
2) ситуации, когда бракованных изделий практически нет.

По результатам исследований сделайте выводы.

Таблица 1. Значения параметров распределений случайной величины.

N	n	p	λ	N	M	a	b
1	20	0.1	1.00	100	90	2	4
2	22	0.11	0.95	110	100	3	5
3	24	0.12	0.90	120	100	2	5
4	26	0.13	0.85	130	120	3	6
5	28	0.14	0.80	140	120	2	5
6	30	0.15	0.75	150	120	3	6
7	21	0.16	0.70	160	150	2	8
8	23	0.17	0.65	170	150	3	9
9	25	0.18	0.60	170	160	2	7
10	27	0.19	0.55	170	130	3	8
11	29	0.20	0.50	170	165	3	10
12	31	0.21	1.05	105	100	4	10
13	20	0.22	1.10	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	1.30	145	120	4	7

ПОРЯДОК ВЫПОЛНЕНИЯ ЗАДАНИЯ

1. Перед началом выполнения задания изучить основной инструментарий Mathcad.
2. Ведите параметры распределения.
3. Определите функцию плотности случайной величины.
4. Определите функцию распределения случайной величины.
5. Постройте графики распределения и функции распределения случайной величины.
6. Определите интервал изменения случайной величины.
7. Найдите по графику наиболее вероятное значение случайной величины.
8. Введите в рабочий документ наибольшее значение вероятности (значение вероятности в точке, вычисленной в предыдущем пункте).
9. Вычислите сумму всех значений вероятностей.
10. Вычислите вероятность попадания значения случайной величины в указанный интервал как разность соответствующих значений функции распределения.
11. Измените значения параметров и повторите вычисления. Сравните полученные результаты, сделайте выводы.
12. Выполните вычисления пп.1÷10 для всех указанных в задании распределений.

СОДЕРЖАНИЕ ОТЧЕТА

1. Цель работы.
2. Задание.
3. Рабочий документ Mathcad для всех указанных распределений с комментариями.
4. Ответы на контрольные вопросы.
5. Выводы.

КОНТРОЛЬНЫЕ ВОПРОСЫ

1. Что называется дискретной случайной величиной? Когда речь может идти о дискретной случайной величине, а когда о непрерывной?
2. Что подразумевают, когда говорят о нормальном распределении?
3. Что называется распределением (рядом распределения) случайной величины?
4. Для чего в рабочем документе Mathcad вводится ORIGIN=1?
5. Как в Mathcad найти по графику наиболее вероятное значение случайной величины?