Лекция 3

Случайные величины

Случайные величины как факторы случайности событий

Среди факторов случайности *события* есть *величины*, значение которых случайно: промах, угол подхода к цели. Все величины, с которыми оперируют в инженерных расчетах, и даже те, что считаются вполне определенными, в реальности под влиянием большого числа неконтролируемых факторов непредсказуемым образом отличаются от своих номинальных значений. Хотя инженерные расчеты выполняют по номинальным значениям, решение по результатам расчетов принимают с запасом, учитывая, что реальные величины (размеры, нагрузки и т.п.) могут иметь случайные отклонения от расчетных.

Определение случайной величины

Случайность величин имеет ту же природу, что и случайность событий. Они различаются лишь математической природой: ***случайная величина – это действительная функция на множестве случайных событий***. Данное определение конструктивно, оно перекидывает мостик между случайными событиями и случайными величинами (СВ). Покажем это сначала на примере дискретных СВ.

Дискретные случайные величины

Дискретные СВ порождаются конечным или счетным множеством случайных событий: $X{:}\{A_1, A_2,\dots, A_n\}\to R^1$. Эту функциональную связь следует понимать так, что случайная величина X принимает одно из своих *возможных значений* $x_i = X(A_i)$, если наступает случайное событие A_i, и это определяет вероятность $p_i = P(X = x_i) = P(A_i)$. СВ обозначаются в тексте большими латинскими буквами, а их возможные значения – малыми.

Свойства дискретных распределений

Так как областью определения СВ является полная группа событий, *сумма вероятностей всех возможных значений равна единице*:

$$\sum_i p_i = \sum_i P(A_i) = P\left(\sum_i A_i\right) = P(\Omega) = 1\,.$$

Это обязательное свойство имеют в виду, когда говорят не просто о вероятностях возможных значений, а о *распределении случайной величины*, то есть о распределении единицы между вероятностями всех ее возможных значений. Таблица, содержащая в одной строке все возможные значения дискретной СВ, *строго упорядоченные по возрастанию*, а в другой – вероятности принятия случайной величиной этих возможных значений, называется *рядом распределения*.

Таблица 3.1. Ряд распределения

$\boldsymbol{x_i}$	x_1	x_2	…	x_k	…
$\boldsymbol{p_i}$	p_1	p_2	…	p_k	…

Индикатор случайного события

Событие A вместе с дополнением $\overline{A}$ составляют полную группу, на которой можно определить функцию $X{:}\{\overline{A}, A\}\to \{0, 1\}$. Она называется *характеристической СВ для события A*, или *индикатором события A*:

$$X = \begin{cases} 1, & \text{если наступает } A, \\ 0, & \text{если наступает } \overline{A}. \end{cases}$$

Очевидно, что $P(X=1) = P(A)$, $P(X=0) = P(\overline{A}) = 1 - P(A)$. Среднее значение характеристической СВ для события A (взвешенное по вероятностям) равно вероятности этого события

$$\sum_i x_i p_i = 0 \cdot P(X=0) + 1 \cdot P(X=1) = P(A).$$

Характеристические СВ позволяют использовать для анализа случайных событий более мощный аппарат случайных величин.

Описание распределений дискретных СВ

Правило, позволяющее находить вероятности любых событий, связанных с данной СВ, называется *законом распределения*. *Ряд распределения* дискретной СВ позволяет найти вероятность любого события, наступлению которого благоприятствуют определенные значения СВ X: $A = (X = x_i)$, $i \in I$. Так как события $(X = x_i)$ несовместны, нужно просто сложить вероятности этих возможных значений: $P(A) = \sum_{i \in I} p_i$.

Поскольку события $(X = x_i)$ составляют полную группу, их можно рассматривать как гипотезы для события A, вероятность которого зависит от x_i. Формулу полной вероятности можно выразить через ряд распределения и условные вероятности $P(A|X = x_i) \equiv P(A|x_i)$:

$$P(A) = \sum_i P(A \mid x_i) p_i. \tag{3.1}$$

Биномиальное распределение

Три независимых выстрела по мишени порождают полную группу события A_0, A_1, A_2, A_3 (событие A_k – произошло ровно k попаданий). СВ X:$\{A_0, A_1, A_2, A_3\} \to R^1$ – число попаданий за стрельбу, ее возможные значения $x_k = X(A_k) = k$, вероятности возможных значений $p_k = P(X = k) = P(A_k) = C_3^k p^k (1-p)^{3-k}$, $k = 0, 1, 2, 3$. Графическое изображение ряда распределения точками (x_i, p_i), соединенными для наглядности отрезками прямых, называется *многоугольником распределения*. Ряд распределения в таблице 3.1 вычислен с помощью функции Ber, этой же командой построен многоугольник распределения (рис. 3.1):

Таблица 3.1. Ряд распределения

x_i	0	1	2	3
p_i	0.216	0.432	0.288	0.064

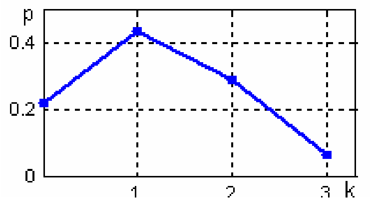


Рис. 3.1. Многоугольник биномиального распределения

```
>> p= Ber(0.4,3),plot(0:3,p)
p =   0.2160    0.4320    0.2880    0.0640
```

Распределение числа успехов в испытаниях Бернулли называется *биномиальным законом распределения*:

$$P(X = k) = p_k = C_n^k p^k (1-p)^{n-k}, \ k = 0, 1, \ldots \tag{3.2}$$

Условия, которым должно удовлетворять распределение, выполнены:

$$p_k > 0, \ \sum_{k=0}^{\infty} C_n^k p^k (1-p)^{n-k} = (p + 1 - p)^n = 1.$$

Распределение Пуассона

Случайная величина X подчиняется закону Пуассона с параметром λ, если она принимает возможные значения из бесконечного ряда неотрицательных целых чисел с вероятностями, задаваемыми формулой Пуассона:

$$P(X = k) = p_k = \frac{\lambda^k}{k!} e^{-\lambda}, \ k = 0, 1, 2, \ldots \tag{3.3}$$

Легко проверить, что сумма членов бесконечного ряда p_k сходится к 1:

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

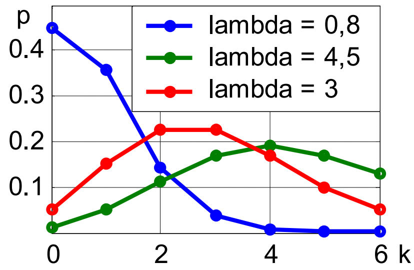


Рис. 3.2. Распределение Пуассона

На рис. 3.2 показаны характерные многоугольники распределения Пуассона с параметрами λ = 0,8, целом λ = 3 и λ = 4,5. Они построены с помощью электронной формулы p_Poisson следующей командой:

```
>> k=0:6;plot(k,p_Poisson(0.8,k),k,p_Poisson(4.5,k),k,p_Poisson(3,k))
```

Геометрическое распределение

В испытаниях Бернулли, которые прекращаются после первого успеха (например, стрельба до первого попадания в цель), число необходимых повторений случайно, его возможные значения – натуральный ряд чисел. Событие $(X = k)$ означает, что первый раз успех наступил в последнем k - м повторении после неудач в предыдущих $k-1$ опытах. Произведение k независимых событий, одно из которых имеет вероятность успеха p, а остальные – вероятность противоположного события $q = 1 - p$, дает k-й элемент ряда распределения:

$$P(X = k) = q^{k-1}p, \ \ k = 1, 2, \ldots \qquad (3.4)$$

Ряд вероятностей образует бесконечно убывающую геометрическую прогрессию с первым членом p и знаменателем q, поэтому

$$\sum_{k=1}^{\infty} q^{k-1}p = \frac{p}{1-q} = \frac{p}{p} = 1 .$$

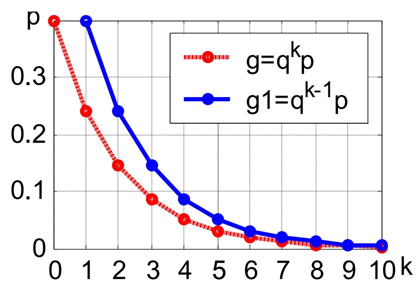


Рис. 3.3. Геометрическое распределение

Распределение (3.4) называется *сдвинутым геометрическим*. *Геометрическим* называется распределение, последовательность возможных значений которого начинается с ноля, а вероятности образуют ту же геометрическую прогрессию, что и в законе (3.4):

$$P(X = k) = q^{k}p, \ \ k = 0, 1, \ldots$$

На рис. 3.3 показаны многоугольники геометрического и сдвинутого геометрического распределений с параметром p = 0,4, построенные следующими командами:

```
>> k=0:10;k1=1:10;p=0.4;q=1-p;g=q.^k*p;g1=q.^(k1-1)*p;
>> plot(k,g,'r--',k1,g1,'b.-'),legend('g=q^kp','g1=q^k^-^1p')
```

Гипергеометрическое распределение

В случайной выборке объема M из партии N изделий, в которой имеется R дефектных, может оказаться случайное количество X дефектных изделий с возможными значениями от 0 до $\min\{M, R\}$. Вероятности возможных значений определяются формулой (1.3):

$$P(X = k) = \frac{C_R^k C_{N-R}^{M-k}}{C_N^M}, \ \ k = 0, 1, \ldots, \min\{M, R\}. \qquad (3.5)$$

Это распределение называется *гипергеометрическим*. Ради вычислительной целесообразности его часто заменяют биномиальным. Действительно, формирование случайной выборки можно рассматривать как испытания Бернулли с числом повторений M и вероятностью успеха $p = R/N$, не меняющейся после каждого случайного выбора при *очень большом* N. Но при малых N результаты случайного выбора по мере уменьшения N уже *нельзя считать равновероятными и независимыми*. С помощью функций Ber и Sampling построим графики биномиального и гипергеометрического распределений для двух вариантов данных с одинаковым относительным числом дефектных изделий в большой (500) и малой (50) партиях:

```
>> N=1000;R=100;M=20;k=0:10;plot(k,Sampling(N,R,M,k),k,Ber(R/N,M,k))
>> N=50;   R=5;                plot(k,Sampling(N,R,M,k),k,Ber(R/N,M,k))
```

Графики на рис. 3.4 подтверждают близость биномиального и гипергеометрического распределений в первом случае. Истинное распределение по закону (3.5) существенно отличается от биномиального, имеющего те же параметры $p = 0{,}1$ и $n = 20$, что и в первом случае. Электронная формула Sampling избавляет от необходимости заменять закон распределения (3.5) биномиальным ради удобства вычислений.

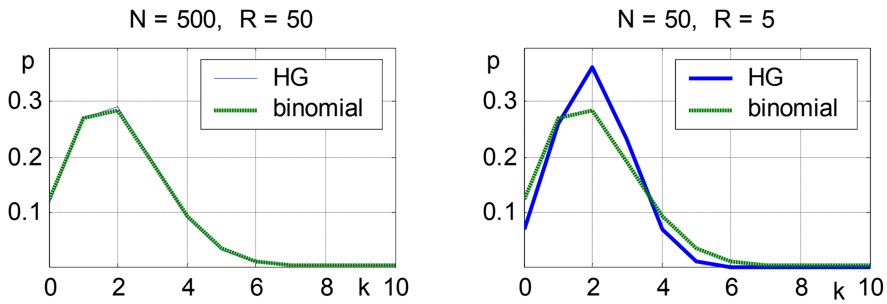


Рис. 3.4. Биномиальное и гипергеометрическое распределения в выборках большого и малого объемов

Непрерывные случайные величины

Вероятностный смысл функции распределения непрерывной СВ

СВ с непрерывной (континуальной) областью возможных значений, таких как угол подхода, нельзя описать распределением вероятностей отдельных значений, так как вероятность любого из них, скорее всего, равна нулю. Определение СВ как функции случайных событий остается в силе, если в качестве аргументов этой функции принимать события $(X < x)$ – «СВ X приняла значение, меньшее данного числа x». Функция $F_X(x) = P(X < x)$ называется *функцией распределения случайной величины X*. Она содержит всю информацию о СВ, в частности, позволяет определить вероятность любого связанного с ней события. Иначе говоря, зная вероятность события $(X < x) = F_X(x)$, можно найти вероятность событий $(X \ge x)$, $(x_1 \le X < x_2)$ и даже $(X = x)$. Действительно, последовательно выразив эти события через событие $(X < x)$

$$(X \ge x) = \overline{(X < x)},$$
$$(x_1 \le X < x_2) = (X < x_2) \setminus (X < x_1),$$
$$(X = x) = \bigcap_{n=1}^{\infty}\left(x \le X < x + \frac{\varepsilon}{n}\right),$$

определим их вероятности через функцию распределения $F_X(x)$:

$$P(X \ge x) = 1 - P(X < x) = 1 - F_X(x),$$
$$P(x_1 \le X < x_2) = P(X < x_2) - P(X < x_2) = F_X(x_2) - F_X(x_1),$$

$$P(X=x)=\lim_{n\to\infty}\mathrm{P}\left(x\le X<x+\frac{\varepsilon}{n}\right)=\lim_{n\to\infty}\left(F\left(x+\frac{\varepsilon}{n}\right)-F(x)\right)=F(x+0)-F(x).$$

Вероятность попадания реализации СВ в интервал равна разности функции распределения на концах интервала, а вероятность значения x равна разности между пределом функции распределения справа от x и ее значением в x.

Свойства функции распределения

Общие свойства функции распределения вытекают из ее вероятностного смысла. Функцией распределения может быть неубывающая функция с областью значений в интервале [0, 1], полунепрерывная слева:

1. $0 \le F(x) \le 1$;
2. $(x_2 \ge x_1) \Rightarrow F(x_2) \ge F(x_1)$;
3. $\lim\limits_{x\to x_0-0} F(x)=F(x_0)$, $\lim\limits_{x\to x_0+0} F(x)=F(x_0)+P(X=x_0)$.

Если возможные значения СВ принадлежат интервалу (a, b), то:

1. $F(x) = 0$ при $x \le a$, так как $P(X < a) = 0$;
2. $F(x) = 1$ при $x > b$, так как $P(X < b + \varepsilon) = 1$.

В любом случае, $\lim\limits_{x\to-\infty} F(x)=0$, $\lim\limits_{x\to\infty} F(x)=1$.

Плотность распределения

СВ называется *непрерывной*, если ее функция распределения непрерывная, кусочно-дифференцируемая функция с непрерывной производной. Для такой СВ можно определить *плотность* (или *плотность вероятности*) распределения как предел отношения вероятности попадания ее значения в бесконечно малый интервал $[x, x+\Delta x)$ к длине этого интервала:

$$f(x)=\lim_{\Delta x\to 0}\frac{P(x\le X<x+\Delta x)}{\Delta x}=\lim_{\Delta x\to 0}\frac{F(x+\Delta x)-F(x)}{\Delta x}=F'(x). \qquad (3.6)$$

В отличие от функции распределения, которая имеет вероятностный смысл для дискретных и непрерывных СВ, плотность существует, когда существует предел в (3.6). Тогда имеет смысл *элемент вероятности* $f(x)dx$ – вероятность попадания СВ на элементарный отрезок $(x, x+dx)$:

$$P(x<X<x+dx) = F(x+dx) - F(x) = \Delta F(x) \approx F'(x)dx = f(x)dx.$$

На графике плотности распределения (рис. 3.5), называемым *кривой распределения*, элемент вероятности – это прямоугольник на отрезке $(x, x+dx)$ с ординатой $f(x)$. Вероятность попадания СВ X в конечный интервал (α, β) равна сумме элементов вероятности на этом интервале, то есть определенному интегралу:

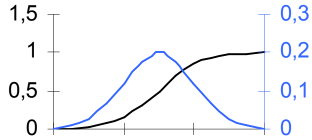


Рис. 3.5. Функция распределения и кривая распределения

$$P(\alpha<X<\beta)=\int_{\alpha}^{\beta} f(x)dx, \qquad (3.7)$$

а в геометрической интерпретации – площади под кривой распределения на данном отрезке оси абсцисс. Функцию распределения можно выразить через плотность $f(x)$ как вероятность попадания в полубесконечный интервал:

$$F(x)=P(X<x)=P(-\infty<X<x)=\int_{-\infty}^{x} f(x)dx. \qquad (3.8)$$

Если функция распределения непрерывна, можно не различать $P(X > \alpha)$ и $P(X \geq \alpha)$, так как вероятность принятия непрерывной СВ любого своего возможного значения нулевая: $P(X = x) = 0$. Но в общем случае эти вероятности отличаются: $P(X \geq \alpha) = P(X > \alpha) + P(X = \alpha)$.

Основное свойство плотности распределения

Как производная неубывающей функции плотность распределения – положительная функция в интервале возможных значений СВ. Основное свойство плотности распределения вытекает из того, что $\lim\limits_{x\to\infty} F(x) = 1$:

$$\int_{-\infty}^{\infty} f(x)dx = 1. \tag{3.9}$$

Это значит, что кривая распределения лежит не ниже оси абсцисс, и площадь под ней равна единице.

Дискретно-непрерывные СВ

Из свойств функции распределения следует, что она может иметь конечное число разрывов первого рода, а в остальных точках диапазона возможных значений монотонно возрастать (рис. 3.6). Это значит, что соответствующая СВ имеет несколько возможных значений x_i с ненулевыми вероятностями p_i, $i\in\{1, 2, \ldots\}$, а в остальных точках ее функция распределения непрерывна. Такие СВ называются *дискретно-непрерывными* или *смешанными*.

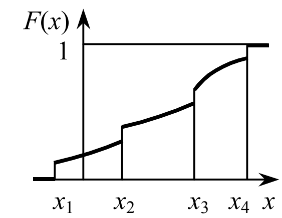


Рис. 3.6. Функция распределения дискретно-непрерывной СВ

Пример построения закона распределения

Длина пробоины в плоском экране от стержневого ПЭ с длиной l, свободно вращающегося при подлете в плоскости, перпендикулярной к экрану, (рис. 3.7) – случайная величина X с возможными значениями в интервале $[0, l]$. Построить функцию распределения $F_X(x) \equiv F(x) = P(X < x)$ значит найти зависимость вероятности события $(X < x)$ от x. Ясно, что $(X < 0)$ – невозможное событие, а длина проекции не превосходит длины стержня, поэтому функция $F(x)$ равна нулю при $x \leq 0$ и единице – при $x > l$. Возможные значения x связаны с длиной отрезка l и углом от нормали φ равенством $x = l \sin \varphi$, следовательно, события $(X < x)$ и $(\varphi < \arcsin(x / l)$) эквивалентны. Так как по условию все угловые положения равновозможны в интервале $[0, \pi/2]$, вероятность события $(X < x)$ можно определить по формуле геометрической вероятности как отношение меры благоприятных исходов $\arcsin(x / l)$ к $\pi / 2$:

$$P(X < x) = P(\varphi < \arcsin(x / l) = \frac{2}{\pi}\arcsin\left(\frac{x}{l}\right).$$

Таким образом, функция и плотность распределения длины пробоины в данных условиях имеют вид:

$$F(x) = \begin{cases} 0, & x < 0; \\ \dfrac{2}{\pi}\arcsin\left(\dfrac{x}{l}\right), & 0 < x < l; \\ 1, & x > l. \end{cases}$$

$$f(x) = F'(x) = \frac{2}{\pi\sqrt{l^2 - x^2}}, \; x \in [0, l].$$

Построим графики функции $f(x)$ в интервале $[0, l)$ без правой границы ($f(x) \to \infty$ при $x \to l$) и функции $F(x)$, непрерывной в интервале, содержащем $[0, l]$ (рис. 3.8):

```
>> L=10;x=0:0.01:1;z=[-0.1, x, 1.1]*L;u=x*L;
>> F=[0,asin(x)*2/pi,1]; f=2/L./(pi*sqrt(1-x(1:end-1).^2));
 >> plot(z,F,u(1:end-1),f),grid
```

Дополним вектор значений $f(x)$ в правом конце интервала так, чтобы выполнялось основное свойство плотности распределения (3.9):

```
>> f(end+1)=(1-Trap(f,u(1:end-1)))*2/0.1-f(end);   Trap(f,u)
ans =     1
```

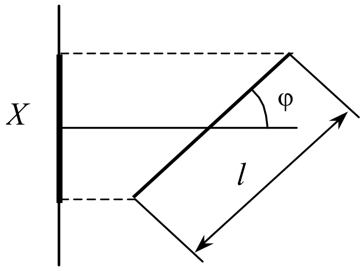


Рис. 3.7.

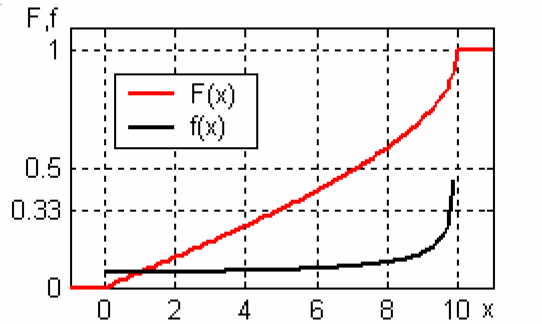


Рис. 3.8.

Интегральный закон распределения длины проекции $F(x)$ дает вероятность того, что эта СВ превысит критическое значение (например, по условию прочности конструкции планера воздушной цели при попадании стержневого ПЭ). Вероятность события $(X > x)$ – это дополнение графика $F(x)$ до единицы. Так, из рис. 3.8 следует, что длина проекции может превысить половину длины стержня с вероятностью 2/3:

$$P(X > 0{,}5l) = 1 - F(0{,}5l) = 1 - \frac{2}{\pi}\arcsin\frac{0{,}5l}{l} = 2/3.$$

Практическое использование законов распределения

Кроме очевидного применения в вычислениях вероятностей попадания в заданный интервал законы распределения представляют СВ во всех операциях: при вычислении среднего значения и других числовых характеристик, анализе стохастического влияния одних СВ на другие и т. д. Вероятностный смысл операций с законами распределения дискретных и непрерывных СВ одинаков. Так, формулу полной вероятности, аналогичную (3.1), можно получить и для непрерывных СВ.

Интегральная формула полной вероятности

Условную вероятность события A в зависимости от значений x непрерывной СВ X обозначают $P(A|x)$, но понимают как $P(A|\,(x < X < x + \Delta x))$, имея в виду, что любое значение в малой окрестности x влияет на вероятность наступления A так же, как x. Для вычисления $P(A)$ можно применить формулу полной вероятности с гипотезами $(x < X < x + dx)$ и их вероятностями $f(x)dx$, заменив суммирование по элементам вероятности интегрированием:

$$P(A) = \int_{-\infty}^{\infty} P(A|x) f(x) dx. \qquad (3.10)$$

Для вычислений по *интегральной формуле полной вероятности* (3.10) удобно использовать электронную формулу Trap (см. Листинг 2.4) так как она позволяет отдельно вычислять подынтегральное выражение на расчетной сетке, а затем выполняет суммирование методом трапеций.

Получение случайных реализаций СВ согласно ее закону распределения

Реализации СВ $X = l \sin \varphi$ можно получить, разыгрывая случайные значения аргумента φ. Равновозможные в интервале $[0, \pi/2]$ значения φ можно получить умножением на $\pi/2$ случайных чисел rand, равномерно распределенных в интервале [0, 1]. Реализации X с помощью датчика случайных чисел можно получить подстановкой $y = \text{rand}$ в функцию $F^{-1}(y) = l \sin(\pi/2\, y)$, обратную к $y = F(x) = (2/\pi) \arcsin(x/l)$. В Лекции 9 будет доказано, что **подстановка случайного числа в функцию, обратную к функции распределения СВ – это общее правило получения реализаций СВ**: $X_i = F^{-1}(\text{rand})$.

Это и так понятно. Во-первых, согласно свойствам функции распределения обратная к ней однозначно отображает интервал [0, 1] на область возможных значений соответствующей СВ. Во-вторых, неравенство rand<y, которое выполняется с вероятностью y, эквивалентно событию $(X < x)$ при $X = F^{-1}(\text{rand})$ и $x = F^{-1}(y)$ в силу монотонности функции F (рис. 3.9), следовательно:

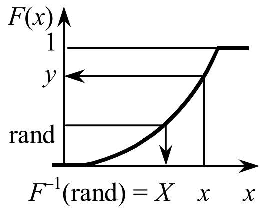


Рис. 3.9. Случайная реализация СВ

$$P(X < x) = P(\text{rand}<y) = y = F(x).$$

Универсальный генератор случайных реализаций СВ

Найти обратное преобразование F^{-1} аналитически можно не для всех распределений. В таких случаях эффективны процедуры использующие вместо математических преобразований вероятностную природу распределения. Для генерации случайного числа, подчиненного биномиальному закону с параметрами n, p, проводят n испытаний Бернулли (n последовательно взятых случайных чисел rand сравнивают с p) и суммируют число успехов, когда rand < p. Реализацию СВ, подчиненной закону Пуассона с параметром λ получают как минимальное количество случайных чисел (без одного), произведение которых меньше, чем $\exp(-\lambda)$, а реализацию геометрического распределения с параметром p – как целую часть величины $\ln(\text{rand}) / \ln(1-p)$.

Функция Gen (Листинг 3.1) возвращает случайные реализации часто употребляемых распределений, сокращенное название которых ('bin' – биномиальное, 'geo' – геометрическое, 'poi' – Пуассона, и т.д.) задано первым аргументом. Вызов Gen с неправильным первым аргументом или без аргументов печатает список всех сокращений для генерируемых распределений. Далее следуют параметры закона в определенном порядке. Порядок аргументов несущественен, если их можно различить по значениям, как, например, параметры биномиального закона $n > 1$, $p < 1$. Последний аргумент задает количество требуемых случайных чисел. Для примера получим 15 реализаций биномиального закона с параметрами $n = 6$, $p = 0{,}4$:

```
>> X=Gen('bin',0.4,6,15),  Y=Gen('bin',6,0.4,15)
X = 2  2  4  4  5  1  2  2  2  3  1  2  4  0  4
Y = 3  2  1  4  1  3  2  0  1  5  1  2  2  3  0
```

Оба выражения записаны правильно, результаты X и Y – различаются как случайные реализации одного и того же распределения Бернулли с вероятностью успеха 0,4.

В отличие от random из библиотеки MATLAB функция Gen умеет работать с объектами, что важно для статистического моделирования попаданий в области различной формы. Например, можно разыграть равномерное распределение в прямоугольнике, круге (фигурах, производных от класса Shape), на их пересечениях и объединениях (рис. 3.10), задав фигуру как единственный параметр:

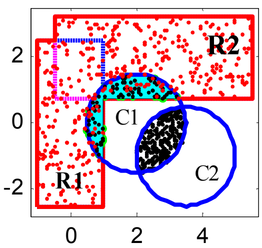


Рис. 3.10. Равномерное распределение точек

```
>>R1=Rect(2,5);R2=Rect(6,2,[3;2]);C1=Circ([2;0],2);C2=C1+[2;-1];
>> Z=Union(R1,R2); S1=Sect(C1,C2); S2=Sect(Z,C1);
>> a=Gen('rnd',Z,500);c1=Gen('rnd',S1,300); c2=Gen('rnd',S2,200);
>> ShowAll(R1,R2,C1,C2,Z,'r',S1,a,'r.',c1,'k.',S2,'Fc',c2,'k.')
```

Статистические распределения

Эмпирическая функция распределения

Закон распределения длины проекции стержня построен на основании гипотезы о равновероятных угловых положениях. Если априорные сведения, необходимые для построения зависимости вероятности $P(X < x)$ от x отсутствуют, но реализации СВ X можно наблюдать в специально поставленном эксперименте (или в явлениях природы), набирают статистику ее реализаций X_j, $j = 1, \ldots, N$ и строят *эмпирическую функцию распределения* $F^*(x)$ по относительной частоте реализаций X, меньших, чем x:

$$F^*(x) = \frac{1}{N}\sum_{X_j<x} 1 .$$

Проведем эксперимент: разыграем $N = 20$ случайных углов в интервале $[0, \pi/2]$ и вычислим соответствующие им длины проекции стержня. В реальном эксперименте распределение случайных углов может быть неизвестным, а длины пробоин измеряются с некоторой случайной погрешностью. Но технология статистической обработки от этого не зависит. Построим график относительных частот с разрывами в реализациях и горизонтальными участками между ними (рис. 3.11):

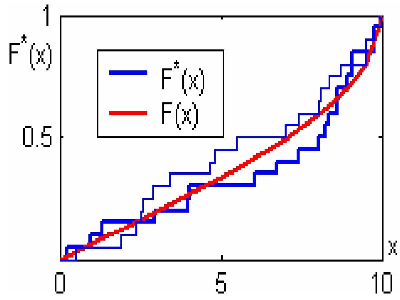


Рис. 3.11. Эмпирическая и теоретическая функции распределения

```
>> L=10;x=0:0.01:1; u=x*L; z=[-0.1, x, 1.1]*L;F=[0,asin(x)*2/pi,1]; f=2/L./(pi*sqrt(1-x(1:end-1).^2));
>> ft=f; N=20; x=rand(1,N)*pi/2;X=sort(sin(x))*L;
>> y=[0 X];y=sort([y,y]);z=(0:N)/N;Z=[z;z];Z=Z(:)';plot(y(2:end),Z(1:end-1), u, F(2:end-1))
```

Если выполнить эти же команды повторно, получим другой график, не совпадающий с первым из-за случайности реализаций при небольшом объеме статистики (две синие линии на рис. 3.11). Для сравнения на тот же график выведена теоретическая функция распределения.

Эмпирический закон распределения дискретной СВ

В случае дискретных СВ можно говорить о частоте реализаций $p_i^* = n_i / N$, где n_i – число реализаций i-о возможного значения, N – объем выборки. По аналогии с рядом распределения – последовательностью пар (x_i, p_i) – и его наглядным представлением в виде многоугольника распределения можно построить *статистический ряд распределения* $(x_i,\ p_i^*)$ и *полигон частот* – соединенные прямыми отрезками точки $(x_i,\ p_i^*)$, $i = 1, \ldots, n$.

Гистограмма частот

Эмпирическая функция распределения *непрерывной* СВ имеет разрывы в случайных точках (см. рис. 3.11). Более наглядное распределение строят на регулярной сетке, для чего область возможных значений делят на интервалы (разряды) h_i, $i = 1, \ldots, n$, по которым распределяют все статистические данные X_j, $j = 1, \ldots, N$. Если $N \gg n$, в каждый разряд h_i попадает достаточное количество n_i экспериментальных точек, чтобы частость $p_i^* = n_i / N$ приближалась к вероятности попадания СВ в i - й интервал. Это условие ограничивает сверху число разрядов при данном N. С другой стороны, разряды должны быть достаточно мелкими, чтобы без больших погрешностей заменить реализации X_j центрами разрядов x_i, в которые они попали. Графическое изображение распределения частот в виде прямоугольников с основаниями на разрядах и высотой, равной частотам, называется *гистограммой*.

Построение гистограммы

Чтобы построить гистограмму длин пробоин от стержня, разыграем больше случайных реализаций (100) в интервале $[0, l]$, разобьем интервал на равные части единичной длины и сгруппируем в них все реализации с помощью функции hist из библиотеки MATLAB, которой нужно передать массив реализаций и центры разрядов:

```
>> L=10; N=100; x=rand(1,N)*pi/2;X=sin(x)*L; h=(1:L)-0.5,m=hist(X,h)
h =  0.50    1.50    2.50    3.50    4.50    5.50    6.50    7.50   8.50    9.50
m =     4    10    10     5     9     8     6    12    12    24
```

Выведем первые 15 из 100 упорядоченных по возрастанию элементов выборки, из них в первый разряд (X<1) попало 4 элемента данных, во второй (1< X <2) – 10, и т.д.:

```
>> Xs=sort(X);Xs(1:15)
ans =  0.1832    0.3181    0.6774    0.7826    1.0119    1.0690    1.2251    1.3016
       1.5679    1.6589    1.6628    1.6992    1.8716    1.9395    2.1813
```

Наложим на гистограмму полученную ранее теоретическую кривую $f(x)$ (рис. 3.12, *а*):

```
>> f1=m/N; bar(h,f1,1,'w'), hold on, plot(u(2:end),f)
```

Так как разряды в данном случае имеют единичную длину, вероятность попадания в каждый из них $f(x)\cdot 1$ равна плотности в центрах интервалов. Большие отклонения частот от точной кривой объясняются недостаточным объемом данных.

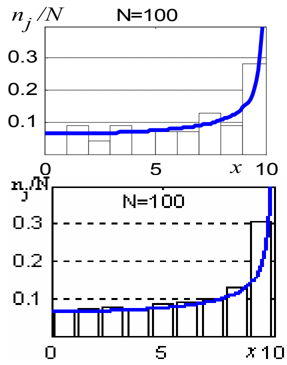


Рис. 3.12. Гистограммы длин проекций стержня

Построитель гистограмм

Улучшить гистограмму можно увеличением объема статистического материала и оптимальным разбиением на разряды. Первый путь затратный, второй способствует снижению затрат или, по крайней мере, лучше их использует. Функция SmartHist (Листинг 3.4) оптимизирует ширину разрядов, если они не заданы. Увеличим число испытаний до 10 000, повторим эксперимент и построим гистограмму с помощью SmartHist в прежних разрядах (рис. 3.12, *б*). Частоты стали близки к вероятностям, а гистограмма (на единичных разрядах) – к плотности распределения:

```
>> X=L*sin(rand(1,10000)*pi/2);[Fs,fs,H]=SmartHist(X,[0 10],10,10);Show(H,'w'),hold on, plot(u(2:end),f)
```

Функция SmartHist возвращает три структуры [F,f,H]. Первые две содержат информацию о статистической функции распределения F и ее производной f (с учетом разрывов). Третья структура содержит разряды H.x и частоты гистограммы H.p, что позволяет использовать ее в автоматическом визуализаторе объектов Show и вычислениях. Например, замена реализаций СВ в разрядах центром разряда превращает среднее арифметическое (случайный результат выборки) в характеристику статистического ряда:

$$m_{\text{ср}} = \frac{1}{N}\sum_{j=1}^{N} X_j \approx \frac{1}{N}\sum_{i=1}^{n} x_i n_i = \sum_{i=1}^{n} x_i \frac{n_i}{N} = \sum_{i=1}^{n} x_i p_i^* \, . \qquad (3.11)$$

Сравним среднее арифметическое и его приближенную оценку (точное среднее значение равно $2l / \pi = 6{,}366$):

```
>> M=mean(X),Mx=dot(H.x, H.p)
M =    6.3938   Mx =    6.3894
```

При достаточно большом объеме статистики сокращение разрядов может улучшить качество гистограммы и вычисленные по ней оценки.

Оптимизация гистограмм

Функция SmartHist оптимизирует количества разрядов и осуществляет оптимальное группирование данных для построения статистической функции и плотности распределения, если не задан третий параметр, она также определяет диапазон возможных значений при отсутствии второго параметра. Построим две гистограммы с разной заданной шириной разрядов и еще одну с оптимальными разрядами (рис. 3.10). Для оценки качества гистограмм выведем также кривые теоретической плотности (пунктирные линии), абсциссы которых пересчитаны на на ожидаемые частоты реализаций в соответствующих разрядах:

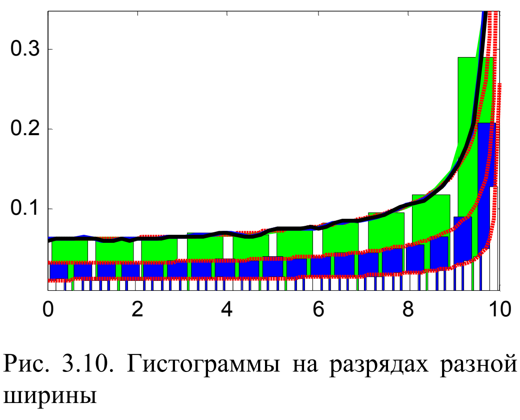


Рис. 3.10. Гистограммы на разрядах разной ширины

```
>> [F,f,H]=SmartHist(X,[0 10],30,10);Show(H,'g',f,'g',[u(2:end);ft*H.h],'r--'), hold on
>> [F,f,H]=SmartHist(X,[0 10],50,20);Show(H,'b',f,'b',[u(2:end);ft*H.h],'r--')
>> [F,f,H]=SmartHist(X);Show(H,'w',f,'k',[u(2:end);ft*H.h], 'r--')
```

Высота прямоугольников гистограммы колеблется вокруг кривой, полученной умножением графика плотности распределения на ширину разряда. На оптимизированной гистограмме (белые столбики) отклонения практически незаметны.

Статистическая функция распределения

Вместо ступенчатой функции строят *статистическую функцию распределения* в виде ломаной с крутизной отрезков, пропорциональной частотам в соответствующих разрядах. Сетка V и сеточные значения Fs выводятся второй парой выходных переменных программы SmartHist. На рис. 3.11 показаны статистическая и эмпирическая функции, построенные по гистограмме (рис. 3.10, *б*) командой:

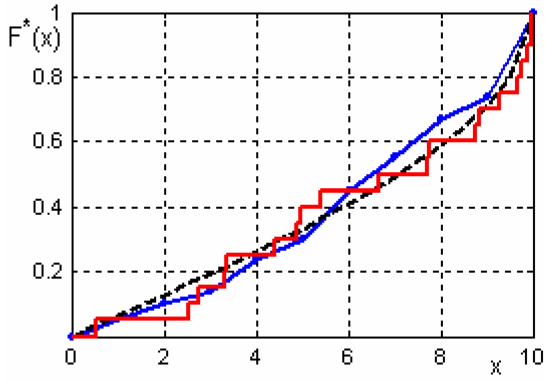


Рис. 3.11. Теоретическая функция распределения и ее статистические приближения

```
>> [v,fs]=SmartHist(X,[0 10],20); plot(v.x,v.F,y(2:end), Z(1:end-1))
```

Функцию SmartHist можно использовать как универсальный построитель гистограмм и статистических распределений, так как она выявляет особенности распределений, имеющиеся в исходном статиститческом материале.

Построение закона распределения дискретной СВ

Построитель гистограмм различает непрерывные, дискретные и дискретно-непрерывные распределения (при достаточном объеме статистического материала). Он выделяет одинаковые значения в нескольких реализациях и по частоте таких реализаций формирует разрывы статистической функции распределения. В качестве примера создадим с помощью Gen массив статистики для СВ, распределенной по биномиальному закону, а затем построим с помощью SmartHist полигон частот и статистическую функцию распределения, которая в данном случае имеет ступенчатый вид. Для сравнения построим и точный многоугольник распределения (рис. 3.14):

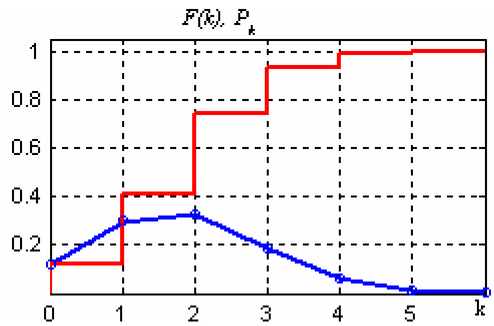


Рис. 3.14. Полигон частот и функция распределения

```
>> [F,f]=SmartHist(Gen('bin',6,0.3,9000),[],20);  Show(f,F,'r', [0:6;Ber(0.3,6)],'o'), grid
```

Моделирование дискретно-непрерывных случайных величин

Пример дискретно-непрерывной СВ

Характерный пример дискретно-непрерывной СВ – площадь перекрытия двух плоских фигур при их случайном взаимном положении.

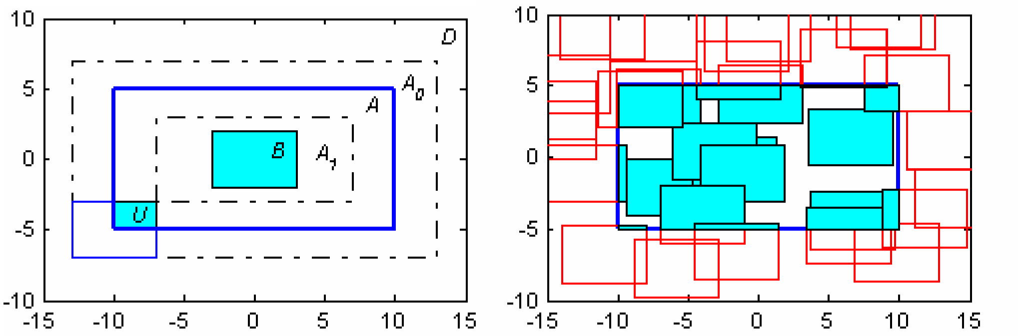
На рис. 3.12 *а* показаны прямоугольник A с размерами 20 × 10 и меньший прямоугольник B со сторонами 6×4. При случайном положении зоны поражения (центр прямоугольника B находится в случайной точке X) площадь перекрытия S также случайна. Построим закон распределения СВ $U = S / S_A$. Возможные значения СВ U принадлежат интервалу $[0, u_m]$, где $u_m = S_B / S_A$, причем $P(U = 0) = p_0 = P(X \notin A_0)$ и $P(U = u_m) = p_1 = P(X \in A_1)$, где A_0 и A_1 – прямоугольники, построенные снаружи и внутри A так, как показано на рис. 3.12 *а*. Внутри интервала $[0, u_m]$ СВ U непрерывна.

Все прямоугольники определим как объекты класса Rect, которые сами могут вычислить свою площадь, занять указанное положение, построить пересечение с другими объектами и сформировать графическое изображение:

```
>> A=Rect(20,10);B=Rect(6,4);b=MySize(B);A1=A-b;A0=A+b;D=Rect(30,20);SA=Area(A)
>> Show(A,'h', A1,'k-.', A0, 'k-.', D, B, 'Fc')
```

В классе Rect функция Sect определяет пересечение двух объектов класса, в результате чего получается объект того же класса (с нулевыми размерами, если пересечение пустое). Прямоугольник в левом нижнем углу рис. 3.15 *а* и пересечение U получены командой:

```
>> Z=moveTo(B,[-10;-5]); U=Sect(A,Z); Show(U,'HFCc',Z,A')
```



а *б*

Рис. 3.15. Прямоугольные цель и зона поражения (*а*), их случайные перекрытия (*б*)

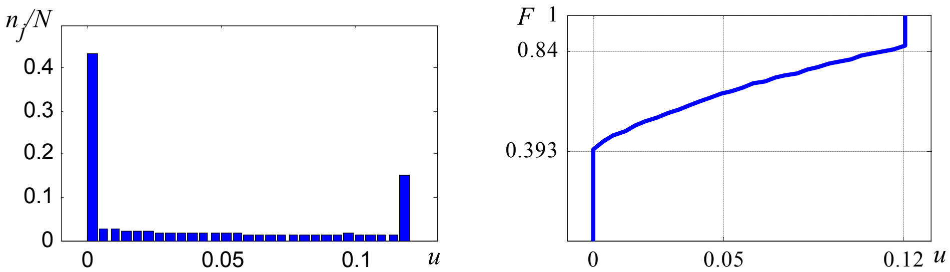
Используем класс Rect в статистическом эксперименте для построения *функции ущерба* $F(u) = P(U < u)$. Распределим случайным образом $N = 1000$ точек в прямоугольнике D, перенесем в эти точки копии зоны B, покажем первые 30 из них, выделив пересечения с A (рис. 3.15 *б*):

```
>> N=10000;X=Gen('rnd',D,N);
>> Z=moveTo(B,X);for i=1:30 T=Sect(A,Z{i});Show(Z{i},'Hr',T,'Fc'); end
```

Теперь вычислим относительную долю накрытия каждой из N зон, построим гистограмму и статистическую функцию распределения:

```
>> S=[];     for i=1:N T=Sect(A,Z{i}); S(i)=Area(T); end
>> S=S/SA; [F,f,H]=SmartHist(S,[],30); Show(H, 501, F, 502)
```

График функции распределения (рис. 3.16, *б*) на концах интервала имеет разрывы, которым соответствуют «всплески» гистограммы. Высота «всплесков» – это частота событий $(U = 0)$ и $(U = u_m)$.



а *б*

Рис. 3.16. Гистограмма относительной доли пораженной площади цели (*а*), статистическая и торетическая функции распределения (*б*)

Полиморфизм объектных методов

Объектная технология замечательна своей полиморфностью, благодаря чему можно повторить вычисления, заменив прямоугольники кругами:

```
>> A=Circ(10);B= Circ(3);D=Circ(15);X=Gen('rnd',D,50);
>> b=MySize(B);A1=A-b;A0=A+b;SA=Area(A);
```

Выполнив остальные команды с новыми объектами A, B, D, построим функцию распределения площадей пересечений. Полиморфное взаимодействие (в данном случае, определение площади пересечений) объектов разных геометрических классов (рис. 3.17) осуществляется корректно каждой парой объектов в соответствии с их свойствами.

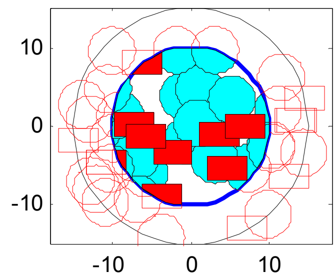


Рис. 3.17. Равномерное распределение фигур

```
>> Show(A,'h', A1,'k-.', A0, 'k-.', D, B, 'Fc')
>> Z=moveTo(B,[-10;-5]); U=Sect(A,Z); Show(U,'HFCc',Z,A')
>> N=100;X=Gen('rnd',D,N);
>> Z=moveTo(B,X);for i=1:30 T=Sect(A,Z{i});Show(Z{i},'Hr',T,'Fc'); end
>> S=[];     for i=1:N T=Sect(A,Z{i}); S(i)=Area(T); end
>> S=S/SA; [F,f,H]=SmartHist(S,[],30); Show(H, 502, F, 503)
```

Программа верификации кода MATLAB

```
clear all
p= Ber(0.4,3),plot(0:3,p)
k=0:6;plot(k,p_Poisson(0.8,k),k,p_Poisson(4.5,k),k,p_Poisson(3,k))
k=0:10;k1=1:10;p=0.4;q=1-p;g=q.^k*p;g1=q.^(k1-1)*p;
plot(k,g,'r--',k1,g1,'b.-'),legend('g=q^kp','g1=q^k^-^1p')
N=1000;R=100;M=20;k=0:10;plot(k,Sampling(N,R,M,k),k,Ber(R/N,M,k))
N=50; R=5; plot(k,Sampling(N,R,M,k),k,Ber(R/N,M,k))

clear all
L=10;x=0:0.01:1;z=[-0.1, x, 1.1]*L;u=x*L;
F=[0,asin(x)*2/pi,1]; f=2/L./(pi*sqrt(1-x(1:end-1).^2));
plot(z,F,u(1:end-1),f),grid
f(end+1)=(1-Trap(f,u(1:end-1)))*2/0.1-f(end);   Trap(f,u)
X=Gen('bin',0.4,6,15),  Y=Gen('bin',6,0.4,15)
R1=Rect(2,5);R2=Rect(6,2,[3;2]);C1=Circ([2;0],2);C2=C1+[2;-1];
Z=Union(R1,R2); S1=Sect(C1,C2); S2=Sect(Z,C1);
a=Gen('rnd',Z,500);c1=Gen('rnd',S1,300); c2=Gen('rnd',S2,200);
ShowAll(R1,R2,C1,C2,Z,'r',S1,a,'r.',c1,'k.',S2,'Fc',c2,'k.')

clear all
L=10;x=0:0.01:1;z=[-0.1, x, 1.1]*L;u=x*L;F=[0,asin(x)*2/pi,1];
f=2/L./(pi*sqrt(1-x(1:end-1).^2));ft=f;
L=10; N=20; x=rand(1,N)*pi/2;X=sort(sin(x))*L;
y=[0 X];y=sort([y,y]);z=(0:N)/N;Z=[z;z];Z=Z(:)';plot(y(2:end),Z(1:end-1), u,
F(2:end-1))
L=10; N=100; x=rand(1,N)*pi/2;X=sin(x)*L; h=(1:L)-0.5,m=hist(X,h)
Xs=sort(X);Xs(1:15)
f1=m/N; bar(h,f1,1,'w'), hold on, plot(u(2:end),ft)
X=L*sin(rand(1,10000)*pi/2);[Fs,fs,H]=SmartHist(X,[0 10],10,10);Show(H,'w'),hold
on, plot(u(2:end),f)
M=mean(X),Mx=dot(H.x, H.p)
[F,f,H]=SmartHist(X,[0 10],30,10);Show(H,'g',f,'g',[u(2:end);ft*H.h],'r--'),
hold on
[F,f,H]=SmartHist(X,[0 10],50,20);Show(H,'b',f,'b',[u(2:end);ft*H.h],'r--')
[F,f,H]=SmartHist(X);Show(H,'w',f,'k',[u(2:end);ft*H.h], 'r--')
[v,fs]=SmartHist(X,[0 10],20); plot(v.x,v.F,y(2:end), Z(1:end-1))
[F,f]=SmartHist(Gen('bin',6,0.3,9000),[],20);  Show(f,F,'r', [0:6;
Ber(0.3,6)],'o')

clear all
A=Rect(20,10);B=Rect(6,4);b=MySize(B);A1=A-b;A0=A+b;D=Rect(30,20);SA=Area(A)
Show(A,'h', A1,'k-.', A0, 'k-.', D, B, 'Fc')
Z=moveTo(B,[-10;-5]); U=Sect(A,Z); Show(U,'HFCc',Z,A')
N=100;X=Gen('rnd',D,N);
Z=moveTo(B,X);for i=1:30 T=Sect(A,Z{i});Show(Z{i},'Hr',T,'Fc'); end
S=[];       for i=1:N T=Sect(A,Z{i}); S(i)=Area(T); end
S=S/SA; [F,f,H]=SmartHist(S,[],30); Show(H, 501, F, 502)
A=Circ(10);B= Circ(3);D=Circ(15);X=Gen('rnd',D,50);

clear all
A=Circ(10);B= Circ(3);D=Circ(15);X=Gen('rnd',D,50);
b=MySize(B);A1=A-b;A0=A+b;SA=Area(A);
Show(A,'h', A1,'k-.', A0, 'k-.', D, B, 'Fc')
Z=moveTo(B,[-10;-5]); U=Sect(A,Z); Show(U,'HFCc',Z,A')
N=100;X=Gen('rnd',D,N);
Z=moveTo(B,X);for i=1:30 T=Sect(A,Z{i});Show(Z{i},'Hr',T,'Fc'); end
S=[];       for i=1:N T=Sect(A,Z{i}); S(i)=Area(T); end
S=S/SA; [F,f,H]=SmartHist(S,[],30); Show(H, 501, F, 502)
```

Контрольные вопросы

1. Объясните связь между случайными событиями и случайными величинами.
2. Какими структурами описывают распределение дискретной СВ?
3. Как отличаются друг от друга многоугольники распределения двух СВ «расход снарядов» и «число промахов» в серии независимых выстрелов до первого попадания?
4. Может ли плотность распределения непрерывной СВ иметь разрывы первого рода?
5. Объясните вероятностный смысл плотности вероятности непрерывной СВ, элемента вероятности. Основное свойство плотности вероятности.
6. Объясните структуру подынтегрального выражения в интегральной формуле полной вероятности.
7. Каковы особенности функции распределения дискретно-непрерывной СВ? Как выполняется основное свойство плотности вероятности дискретно-непрерывных СВ?
8. Как записать интегральную формулу полной вероятности для дискретно-непрерывной СВ? Объясните вероятностный смысл подынтегрального выражения в интегральной формуле полной вероятности.
9. Как построить гистограмму статистического распределения? какими соображениями следует руководствоваться при выборе ширины регистров?
10. Как построить статистический ряд распределения и полигон частот?