Галанов Ю.И.

Лабораторный практикум по мат. статистике

Моделирование случайных величин и их распределений

Моделирование случайных событий.Схема Бернулли

Опыт с двумя исходами полностью определяется заданием вероятности ***p*** события ***A.*** Используя понятие геометрической вероятности, легко показать, что события {**A**} и {**rnd(1) < p**} равносильны.

Для генерирования случайных чисел, равномерно распределенных на отрезке [0,1] используем встроенную функцию ***rnd(1)*.**

Зададим число ***p,*** принадлежащее отрезку [0,1].

$$p := 0.75$$

Будем считать, что появление случайного числа

$$rnd(1) \le p$$

соответствует появлению события ***A.***

Введем индикатор событи ***J(p)***,равный единице, если событие произошло (т.е.если ***rnd(1)<=p***) и равный нулю, если событие не произошло (т.е. ***rnd(1)> p***)

$$J(p) := if(rnd(1) \le p, 1, 0)$$

Выборка из биномиального распределения

Смоделируем опыт, состоящий из ***n*** независимых испытаний и подсчитаем число испытаний, в которых произошло событие ***A*** в данной серии. Это число равно сумме индикаторов всех испытаний в одной серии:

$$n := 15 \qquad i := 0..n \qquad x_i := J(0.25)$$

$$x^T = (0\ \ 0\ \ 0\ \ 0\ \ 1\ \ 0\ \ 1\ \ 0\ \ 0\ \ 0\ \ 0\ \ 1\ \ 0\ \ 1\ \ 0\ \ 1) \qquad \sum x = 5$$

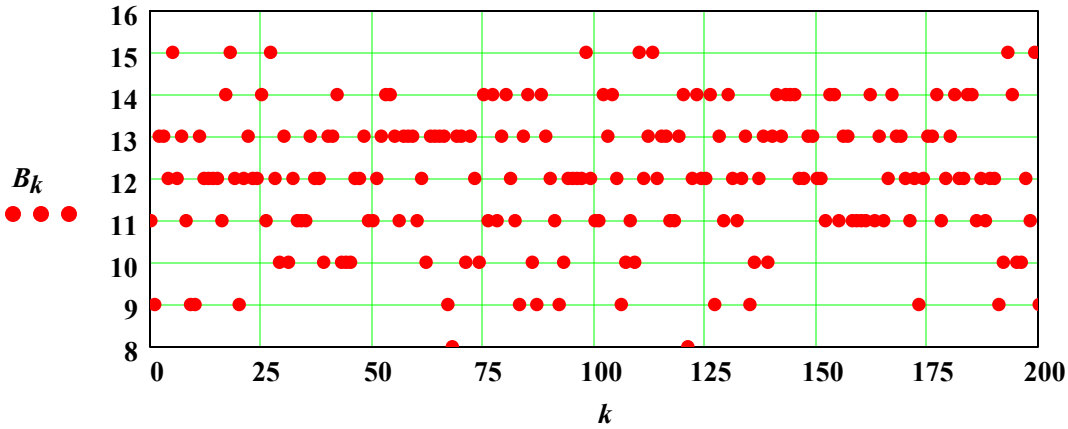
Число появлений события ***A*** в ***n*** независимых испытаниях ***B*** является случайной величиной, подчиняющейся биномиаальному распределению.

Создадим выборку ***B***, состоящую из ***N*** элементов.

$$N := 201 \qquad k := 0 .. N-1 \qquad B_k := \sum_i J(p)$$

Отображение выборки

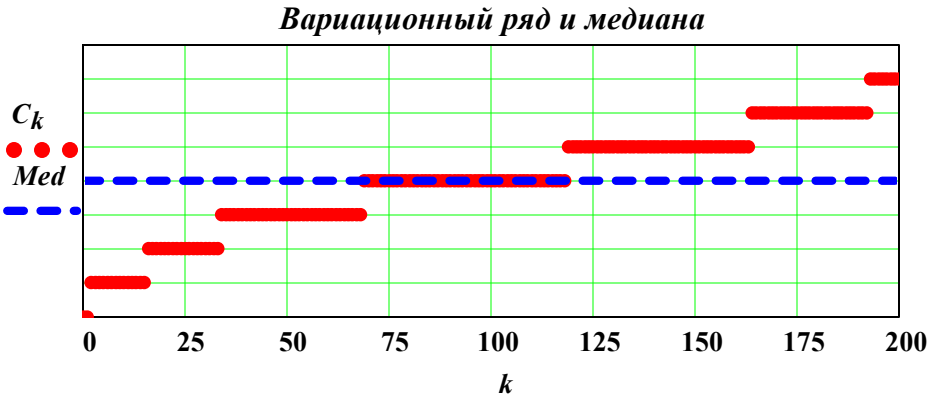
1.РЯД РАСПРЕДЕЛЕНИЯ



2. ВАРИАЦИОННЫЙ РЯД

Отсортируем выборку и получим вариационный ряд, из которого легко найти выборочную медиану и функцию распределения:

$$C := sort(B) \qquad Med := C_{\frac{N-1}{2}} \qquad Med = 12$$



<=...медиана

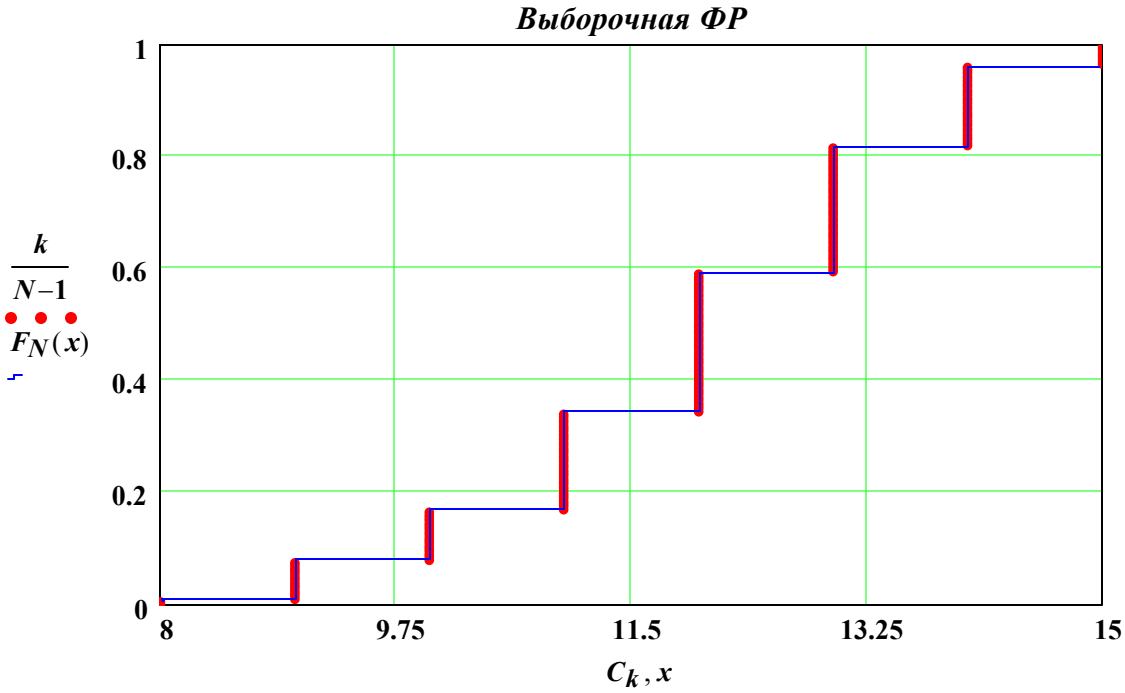
3.ВЫБОРОЧНАЯ ФУНКЦИЯ РАСПРЕДЕЛЕНИЯ

Выборочную функцию распределения можно получить двумя способами:

1) задав ее в явном виде:

$$x := min(B)..n \qquad F_N(x) := \frac{1}{N}\cdot\sum_k if\left(B_k \le x, 1, 0\right) \qquad F_N(80) = 1$$

2)или используя вариационный ряд: => $F_N(k) := \frac{k}{N}$



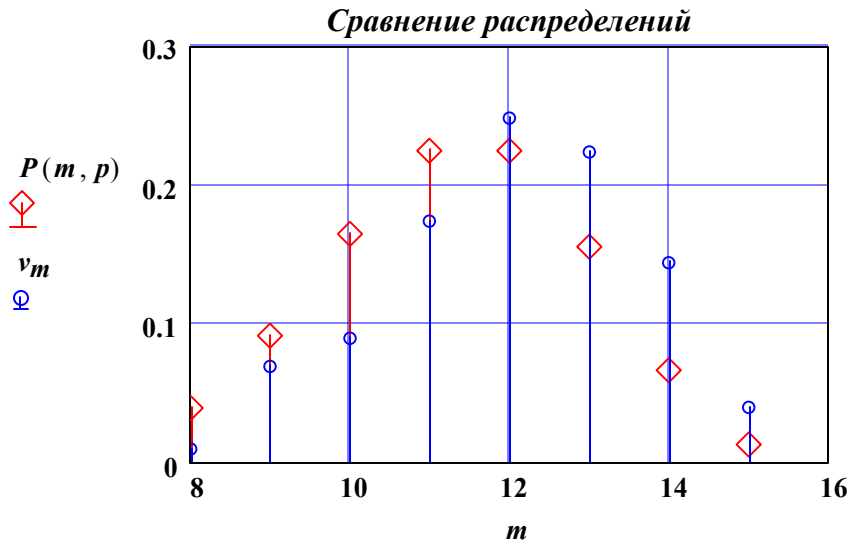
4.Сравним выборочное частотное распределение (статистический ряд) и теоретическое, рассчитанное по формуле Бернулли:

$$P(m, p) := \frac{n!\cdot p^m\cdot(1-p)^{n-m}}{m!\cdot(n-m)!}$$

Вводим индикатор равенства двух чисел.
Затем для каждого значения m подсчитываем, сколько элементов выборки принимает данное значение

$$Ident(x, y) := if(x \neq y, 0, 1) \qquad m := C_0 .. C_{last(C)}$$

$$v_m := \frac{1}{N} \cdot \left(\sum_k Ident(m, C_k) \right)$$



5. ГИСТОГРАММА И ПОЛИГОН ЧАСТОТ

В случае *дискретных* случайных величин строятся их частотные распределения. Гистограмма и полигон частот служат оценками плотности распределения для *непрерывных* случайных величин. В данном примере мы рассмотрим порядок построения гистограммы и полигона частот с использованием функции ***hist***, встроеенной в MathCad, которая автоматически подсчитывает число точек, попавших в каждый полуинтервал.

1. Зададим полуинтервал ***D***, полностью покрывающий область значений выборки (размах выборки) [***a***,***b***]. Разоб"ем его на ***M*** одинаковых полуинтервалов $[d_i, d_{i+1})$:
так,чтобы выполнялось условие $a < d_0 < d_1 < d_2 < \blacksquare .. b < d_M \blacksquare$

Число интервалов группировки зададип по формуле Старджеса:

$$M := 1 + trunc(log(n, 2)) \qquad l := 0 .. M$$

2. Зададим шаг дискретизации

$$h := \frac{max(B) - min(B)}{M} \cdot 1.0001$$

3. Создадим массив ***d***, содержащий координаты точек разбиения области ***D:***

$$d_l := min(B) + h \cdot l \qquad min(B) = 8 \qquad max(B) = 15$$

$$d^T = (8 \quad 9.75 \quad 11.5 \quad 13.251 \quad 15.001)$$

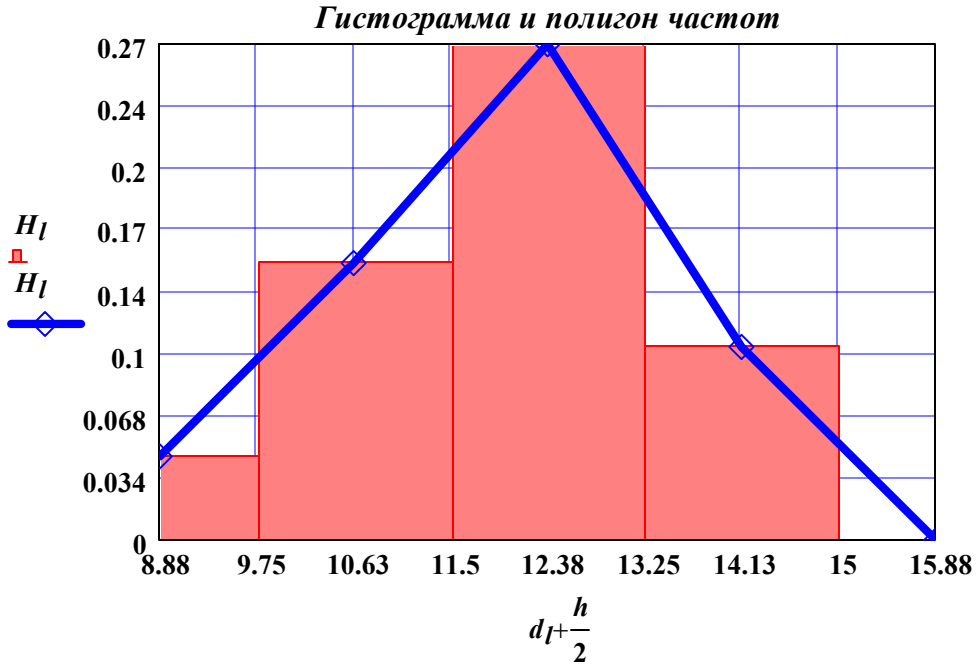
При построении гистограммы используем системную функцию ***hist(vx,vy)***:

$$H := \frac{hist(d, B)}{N \cdot h} \qquad H_{last(H)+1} := 0 \qquad \sum_l H_l \cdot N \cdot h = 201$$ <= Проверочная сумма

Поскольку число элементов у массива ***H*** на единицу меньше,чем у массива ***d***,то мы добавли ему еще один нулевой элемент. При построении графиков в качестве абсцисс берем середины частичных полуинтервалов.

Гистограмма изображается в виде прямоугольников с высотой, равной значению оценки плотности распределения.

Полигон частот - соединяет точки отрезками прямых линий.



6. Оценим выборчные параметры и сравним их с теоретическими (модельными)

- математическое ожидание и выборочное среднее

$$m := n \cdot p \quad m = 11.25 \qquad MO := \frac{1}{N} \cdot \sum_k C_k \quad MO = 12.03$$

- дисперсия и выборочная дисперсия

$$Disp := n \cdot p \cdot (1 - p) \qquad Disp = 2.813 \quad S2 := \frac{1}{N} \cdot \sum_k \left(B_k\right)^2 - MO^2 \quad S2 = 2.477$$

- исправленная дисперсия: $\frac{N}{N-1} \cdot S2 = 2.489$