ЛЕКЦИЯ 3

Дискретные случайные величины. Биномиальное распределение дискретной случайной величины. Числовые характеристики дискретных случайных величин. Системы случайных величин, их числовые характеристики.

1.8. ДИСКРЕТНЫЕ СЛУЧАЙНЫЕ ВЕЛИЧИНЫ

Если результатом эксперимента является число, значение которого нельзя предсказать точно до проведения самого эксперимента, то это число называют *случайной величиной*. Случайную величину называют *дискретной*, если она принимает отдельные, изолированные возможные значения с определенными вероятностями. Число возможных значений дискретной случайной величины может быть конечным или счетным.

Рассмотрим случайную величину X, возможные значения которой образуют конечную или бесконечную последовательность чисел x_1, x_2, ..., x_N ,… . Пусть задана функция $p(x)$, значение которой в каждой точке $x = x_i$ (i =1, 2, …) равно вероятности того, что величина X примет значение x_i

$$p(x_i) = p_i = P\,(X = x_i).$$

. Функция $p(x)$ называется законом распределения вероятностей случайной величины, или кратко, законом распределения. Эта функция определена в точках последовательности x_1, x_2, ..., x_N ,…и обычно задается таблицей вида:

X	x_1	x_2	…	x_N
P	p_1	p_2	…	p_N

где x_i упорядочены по возрастанию $x_1 < x_2 < x_3 < ... < x_N < …$.Так как события $X = x_i$ образуют полную группу, то

$$\sum_{i=1}^{N} p_i = 1,$$ в случае конечной последовательности чисел x_i ,

$\sum_{i=1}^{\infty} p_i = 1$, в случае бесконечной последовательности чисел x_i ,

что часто служит контролем вычисления значений p_i.

Функция распределения вероятностей случайной величины и ее свойства.

Рассмотрим функцию $F(x)$, определенную на всей числовой оси следующим образом: для каждого x значение $F(x)$ равно вероятности того, что дискретная случайная величина примет значение, меньшее x, т. е.

$$F(x) = P(X < x). \tag{1.17}$$

Эта функция называется *функцией распределения вероятностей*, или кратко, *функцией распределения.*

Зная функцию распределения $F(x)$, легко найти вероятность того, что случайная величина X удовлетворяет неравенствам $x_1 \le X < x_2$.

Рассмотрим событие, заключающееся в том, что случайная величина примет значение, меньшее x_2. Это событие распадается на сумму двух несовместных событий: 1) случайная величина X принимает значения, меньшие x_1, т. е. $X < x_1$; 2) случайная величина X принимает значения, удовлетворяющие неравенствам $x_1 \le X < x_2$. Используя правило сложения вероятностей, получаем

$$P(X < x_2) = P(X < x_1) + \mathrm{P}(x_1 \le X < x_2).$$

Отсюда

$$P(x_1 \le X < x_2) = P(X < x_2) - P(X < x_1).$$

Но по определению функции распределения $F(x)$ [см. формулу (1.17)], имеем

$$P(X < x_2) = F(x_2), \quad P(X < x_1) = F(x_1);$$

следовательно,

$$P(x_1 \le X < x_2) = F(x_2) - F(x_1) \tag{1.18}$$

Таким образом, *вероятность попадания дискретной случайной величины в промежуток* $x_1 \le X < x_2$ *равна приращению функции распределения на этом интервале*.

Рассмотрим основные свойства функции распределения.

1°. Функция распределения является неубывающей.

В самом деле, пусть $x_1 < x_2$. Так как вероятность любого события неотрицательна, то $P(x_1 \le X < x_2) \ge 0$. Поэтому из формулы (1.18) следует, что $F(x_2) - F(x_1) \ge 0$, т. е.

$F(x_2) \ge F(x_1)$.

2°. Значения функции распределения удовлетворяют неравенствам $0 \le F(x) \le 1$.

Это свойство вытекает из того, что $F(x)$ определяется как вероятность [см. формулу (1.17)]. Ясно, что $F(-\infty) = 0$ и $F(+\infty) = 1$.

3°. Вероятность того, что дискретная случайная величина примет одно из возможных значений x_i, равна скачку функции распределения в точке x_i.

Действительно, пусть x_i – значение, принимаемое дискретной случайной величиной, и $\Delta x > 0$. Полагая в формуле (1.18) $x_1 = x_i$, $x_2 = x_i + \Delta x$, получим

$$P(x_i \le X < x_i + \Delta x) = F(x_i + \Delta x) - F(x_i) \qquad (1.19)$$

В пределе при $\Delta x \to 0$ вместо вероятности попадания случайной величины на интервал $x_i \le X < x_i + \Delta x$ получим вероятность того, что величина X примет данное значение x_i:

$$\lim_{\Delta x \to 0} P(x_i \le X < x_i + \Delta x) = P(X = x_i) = p(x_i).$$

С другой стороны, получаем $\lim\limits_{\Delta x \to 0} F(x_i + \Delta x) = F(x_i + 0)$, т. е. предел функции $F(x)$ справа, так как $\Delta x > 0$. Следовательно, в пределе формула (1.19) примет вид

$$p(x_i) = F(x_i + 0) - F(x_i) = F(x_i + 0) - F(x_i - 0),$$

т. е. значение $p(x_i)$ равно скачку функции в точке x_i

Зная закон распределения дискретной случайной величины X, можно вычислить функцию распределения, представляющую собой, в силу определения (1.13) функцию накопленных вероятностей:

$$F(x) = \sum_{x_i < x} P(X = x_i)\,, \qquad (1.20)$$

где суммирование распространяется на все значения индекса i, для которых $x_i < x$.

1.9. БИНОМИАЛЬНОЕ РАСПРЕДЕЛЕНИЕ ДИСКРЕТНОЙ СЛУЧАЙНОЙ ВЕЛИЧИНЫ.

Пусть проводится n независимых испытаний, в каждом из которых событие A наступает с вероятностью p ($0 < p < 1$), и не наступает с вероятностью q, $q = 1 - p$. Обозначим $P_n(m)$ – вероятность того, что в n испытаниях событие A наступит ровно m раз. При этом заметим, что наступления или ненаступления события A могут чередоваться различным образом. Условимся записывать возможные результаты испытаний в виде комбинаций букв A и $\overline{A}$. Например, запись $A\overline{A}\,\overline{A}A$ означает, что в четырех испытаниях событие осуществилось в 1-м и 4-м случаях и не осуществилось во 2-м и 3-м случаях.

Всякую комбинацию, в которую A входит m раз и $\overline{A}$ входит $n - m$ раз, назовем благоприятной. Количество благоприятных комбинаций равно количеству k способов, которыми можно выбрать m чисел из данных n; таким образом, оно равно числу сочетаний из n элементов по m, т. е. $k = C_n^m$.

Подсчитаем вероятности благоприятных комбинаций. Рассмотрим сначала случай, когда событие A происходит в первых m испытаниях и, следовательно, не происходит в остальных $n - m$ испытаниях. Такая благоприятная комбинация имеет следующий вид:

$$B_1 = AA...A\,\overline{A}\overline{A}...\overline{A},$$

где событие A повторяется m раз, а затем событие $\overline{A}$ повторяется $n - m$ раз. Вероятность этой комбинации в силу независимости испытаний (на основании теоремы умножения вероятностей) составляет

$$P(B_1) = P(A)P(A)...P(A)\;P(\overline{A})P(\overline{A})...P(\overline{A}) = p^m q^{n-m}.$$

Так как в любой другой благоприятной комбинации B_i событие A встречается также m раз, а событие $\overline{A}$ происходит $n - m$ раз, то вероятность каждой из таких комбинаций также равна $p^m q^{n-m}$. Итак,

$$P(B_1) = P(B_2) = \ldots = P(B_k) = p^m q^{n-m}.$$

Все благоприятные комбинации являются, очевидно, несовместными. Поэтому (на основании правила сложения вероятностей)

$$P_n(m) = P(B_1 + B_2 + \ldots + B_k) = P(B_1) + P(B_2) + \ldots + P(B_k) = kp^m q^{n-m} = C_n^m p^m q^{n-m}.$$

Следовательно, вероятности $P_n(m)$ вычисляются по формуле:

$$P_n(m) = C_n^m p^m q^{n-m}, \tag{1.21}$$

где коэффициенты C_n^m называются числом сочетаний из n элементов по m и вычисляют по формуле:

$$C_n^m = \frac{n!}{m!(n-m)!} = \frac{n(n-1)(n-2)\ldots(n-m+1)}{1\cdot 2\cdot 3\ldots m}. \tag{1.22}$$

тогда случайная величина X такая, что $P(X=m)= P_n(m)$ определяет *биномиальное распределение* или *распределение Бернулли*.

Из формулы (1.22) видно, что $C_n^m = C_n^{n-m}$ $(m = 1, 2, \ldots, n-1)$. По определению полагают также $C_n^0 = C_n^n = 1$.

Рассмотрим еще раз смысл коэффициентов C_n^m. Пусть заданы n разных элементов. Всевозможные группировки из данных n элементов по m элементов в каждой, отличающиеся друг от друга хотя бы одним элементом, при этом порядок расположения элементов в группировке безразличен, называются *сочетаниями* из n элементов по m. Например, $n = 4$, имеем 4 элемента: *a*, *b*, *c*, *d*. Выпишем сочетания из четырех элементов по два: *ab*, *ac*, *ad*, *bc*, *bd*, *cd*. Из определения следует, что сочетания *ab* и *ba* не различимы. Число таких сочетаний и находится по формуле (1.18), $C_4^2 = \dfrac{4\cdot 3}{1\cdot 2} = 6$.

Примеры вычисления: $C_9^3 = \dfrac{9\cdot 8\cdot 7}{1\cdot 2\cdot 3} = 84$; $C_{18}^{16} = C_{18}^2 = \dfrac{18\cdot 17}{1\cdot 2} = 153.$

Вероятность того, что в n испытаниях событие A наступит менее m раз, равна сумме вероятностей $P_n(0) + P_n(1) + ... + P_n(m-1)$; более m раз – сумме вероятностей $P_n(m+1) + ... + P_n(n)$. Часто расчеты упрощаются, если применять свойство вероятностей $P_n(m)$

$$P_n(0) + P_n(1) + \ldots + P_n(n) = 1.$$

Часто необходимо знать, при каком значении m вероятность $P(X=m)$ для случайной величина X, имеющей распределение Бернулли, .принимает наибольшее значение. То есть требуется найти *наивероятнейшее число* m_0 *наступления события* A в данной серии опытов. Можно доказать, что число m_0 должно удовлетворять двойному неравенству

$$np - q \leq m_0 \leq np + p. \qquad (1.23)$$

Заметим, что сегмент $[np - q,\ np + p]$, в котором лежит m_0, имеет длину

$$(np+p) - (np - q) = p+q=1.$$

Поэтому, если какой-либо из его концов не является целым числом, то между этими концами лежит единственное целое число, и m_0 определено однозначно. В том случае, если оба конца – целые числа, имеются два наивероятнейших значения: $np - q$ и $np + p$.

При больших значениях n подсчет вероятностей $P_n(m)$ по формуле (1.17) связан с громоздкими вычислениями. В этом случае удобнее пользоваться следующей формулой:

$$P_n(m) \approx \frac{1}{\sqrt{npq}\sqrt{2\pi}} e^{-x^2/2} = \frac{1}{\sqrt{npq}} \varphi(x) \qquad (1.24)$$

где $x = \dfrac{m - np}{\sqrt{npq}}$ (p не равно нулю и единице), а $\varphi(x) = \dfrac{1}{\sqrt{2\pi}} e^{-x^2/2}$.

Формула (1.20) выражает *локальную теорему Лапласа*. Точность этой формулы повышается с возрастанием *n*. Функция $\varphi(x)$, как мы увидим в дальнейшем, играет очень большую роль в теории вероятностей.

На практике часто встречается случай биномиального распределения, при котором n велико, а p мало. Например, X – число бракованных изделий объема n (n – велико) при вероятности появления бракованного изделия $p << 1$.

Этот случай можно рассматривать как *асимптотику биномиального распределения* при $n \to \infty$, $p \to 0$, $np = \lambda = Const$.

Представим выражение для вероятности в биномиальном распределении в виде

$$P(X = m) = \frac{1}{m!}(np)(np - p)\ldots(np - (m-1)p)\cdot(1-p)^{n-m}.$$

Перейдем к пределу:

$$P(X = m) = \frac{1}{m!}\lambda^m \lim_{n\to\infty}\left(1 - \frac{\lambda}{n}\right)^{n-m} = \frac{\lambda^m}{m!}e^{-\lambda}; \quad (m = 0,1,2,\ldots). \qquad (1.25)$$

Полученное распределение называется *распределением Пуассона*. Распределение Пуассона с достаточной точностью может быть использовано как приближенное для биномиального распределения при $n \geq 100$, $p \leq 0{,}01$.

1.10. ЧИСЛОВЫЕ ХАРАКТЕРИСТИКИ ДИСКРЕТНЫХ СЛУЧАЙНЫХ ВЕЛИЧИН

Случайные величины могут описываться числовыми характеристиками, среди которых различают характеристики положения (математическое ожидание, мода, медиана) и характеристики рассеяния (дисперсия, среднеквадратическое отклонение).

Пусть X – дискретная случайная величина с заданным законом распределения $P(X=x_i) = p_i$ ($i = 1, 2,\ldots, N$; $\sum_{i=1}^{N} p_i = 1$). *Математическое ожидание* $M(X)$ представляет собой среднее ожидаемое значение случайной величины. Эта величина вычисляется по формуле:

$$M(X) = \sum_{i=1}^{N} x_i p_i\,, \qquad (1.26)$$

Свойства математического ожидания:

1) $M(C) = C$, где $C = const$,

2) $M(CX) = CM(X)$, (1.27)

3) $M(X+Y) = M(X) + M(Y)$ для любых случайных величин X и Y,

4) $M(XY) = M(X) \cdot M(Y)$ если X и Y независимы.

Таким образом, если случайная величина X представлена в виде линейной комбинации величин X_1, X_2, ..., X_L, то ее математическое ожидание вычисляют ,пользуясь свойством линейности:

$$M(X) = M\left(\sum_{i=1}^{L} c_i X_i\right) = \sum_{i=1}^{L} c_i M(X_i). \qquad (1.28)$$

Если известен закон распределения дискретной случайной величины, то математическое ожидание любой ее функции может быть вычислено по формуле:

$$M(f(X)) = \sum_{i=1}^{N} f(x_i) \cdot p_i. \qquad (1.29)$$

Для характеристики степени отклонения значений случайной величины от ее математического ожидания $M(X) = a$ вводятся понятия *дисперсии* $D(X)$ и *среднего квадратического отклонения* $\sigma(X)$ по формулам:

$$D(X) = M(X-a)^2; \qquad \sigma(X) = \sqrt{D(X)} \qquad (1.30)$$

Свойства дисперсии и среднего квадратического отклонения:

1) $D(C) = 0$; $\sigma(C) = 0$,

2) $D(X+C) = D(X)$; (1.31)

3) $D(CX) = C^2 D(X)$; $\sigma(CX) = |C|\sigma(X)$,

4) если X и Y независимы, то $D(X+Y) = D(X) + D(Y)$, .

где $C = const$

Можно доказать, что дисперсия для дискретной случайной величины X может быть найдена по формуле:

$$D(X) = M(X^2) - (M(X))^2 = \sum_{i=1}^{N} x_i^2 p_i - (M(X))^2. \qquad (1.32)$$

Математическое ожидание и дисперсия дискретной случайной величины X, имеющей *биномиальное распределение*, могут быть найдены по формулам:

$$M(X) = np; \quad D(X) = npq\ , \tag{1.33}$$

где p – вероятность того, что событие A произойдет, q – вероятность того, что событие A не произойдет в каждом из независимых испытаний.

Для распределения Пуассона $M(X) = D(X) = \lambda$.

1.11. СИСТЕМЫ СЛУЧАЙНЫХ ВЕЛИЧИН, ИХ ЧИСЛОВЫЕ ХАРАКТЕРИСТИКИ.

Две или несколько случайных величин $X_1, X_2, ..., X_L$ образуют *систему случайных величин*. Система случайных величин может рассматриваться также как *случайный вектор* $\vec{X} = (X_1, X_2, ..., X_L)^T$.

Основными характеристиками системы случайных величин служат:

центр распределения $$M = \begin{pmatrix} M(X_1) \\ M(X_2) \\ \dots \\ M(X_L) \end{pmatrix} = \begin{pmatrix} a_1 \\ a_2 \\ \dots \\ a_L \end{pmatrix}$$

и *матрица ковариаций*

$$\mathrm{K} = (K_{ij}) = \begin{pmatrix} K_{11} & K_{12} & \dots & K_{1L} \\ K_{21} & K_{22} & \dots & K_{2L} \\ \dots & \dots & \dots & \dots \\ K_{L1} & K_{L2} & \dots & K_{LL} \end{pmatrix} \tag{1.34}$$

где $a_i = M(X_i)$ – математическое ожидание величины X_i ($i = 1, 2, ..., L$);

$K_{ij} = K_{ji}$ – *ковариация* между величинами X_i и X_j, ($i, j = 1, 2, ..., L$, $i \neq j$).

$$K_{ij} = \mathrm{cov}(X_i X_j) = M[(X_i - a_i)(X_j - a_j)], \tag{1.35}$$

$K_{ii} = D(X_i)$ – дисперсия величины X_i.

Ковариация между X_i и X_j для дискретных случайных величин может быть найдена по формуле:

$$K_{ij} = M(X_i X_j) - a_i a_j = \sum_{k=1}^{N} x_{ik} x_{jk} p_k - a_i a_j \,. \qquad (1.36)$$

Две случайные величины *X* и *Y* являются (статистически) независимыми, если для любых функций $f(x)$ и $g(x)$, для которых существуют математические ожидания, имеет место соотношение

$$M[f(x) \cdot g(x)] = M[f(x)] \cdot M[g(x)] \,. \qquad (1.37)$$

Смысл этого определения состоит в том, что никакая информация о значениях одной величины не влияет на любую информацию о значениях другой величины. В частности, из (1.37) следует

$$M(X \cdot Y) = M(X) \cdot M(Y), \qquad (1.38)$$

а также тот факт, что из независимости двух случайных величин следует их некоррелированность, т.е., если *X* и *Y* независимы, то $\mathrm{cov}(X,Y) = 0$.

Дисперсия линейной комбинации случайных величин выражается через ковариации:

$$D\left(\sum_{i=1}^{L} c_i X_i\right) = \sum_{i=1}^{L} \sum_{j=1}^{L} c_i c_j K_{ij} \,. \qquad (1.39)$$

В частности, для попарно независимых величин X_i и X_j ковариации $K_{ij} = 0$ ($i, j = 1, 2, ..., L, i \neq j$) и матрица ковариаций диагональна, а дисперсия линейной комбинации величин выражается только через их дисперсии:

$$D\left(\sum_{i=1}^{L} c_i X_i\right) = \sum_{i=1}^{L} c_i^2 D(X_i) \,. \qquad (1.40)$$

Ковариация $K(X,Y)$ характеризует степень зависимости случайных величин *X* и *Y*. Величина $K(X,Y)$ зависит от единиц измерения случайных величин *X* и *Y*. Во избежании этого неудобства часто вычисляют *коэффициент корреляции:*

$$r(X,Y)=\frac{K(X,Y)}{\sigma(X)\cdot\sigma(\mathrm{Y})}.$$

Коэффициент корреляции – безразмерная величина. Можно доказать, что $|r(X,Y)|\le 1$ для любых X и Y, причем $|r(X,Y)|=1$ тогда и только тогда, когда X и Y связаны линейной зависимостью: $Y=aX+b$, где $a,b\in R,\quad a\neq 0$. $r(X,Y)=1$ если $a>0$ и $r(X,Y)=-1$, если $a<0$. Чем ближе $r(X,Y)$ к единице, тем вероятнее линейная зависимость между случайными величинами X и Y.