Лекция 7. Основные понятия математической статистики

Предмет математической статистики. Генеральная совокупность и выборка. Эмпирическая функция распределения и гистограмма. Числовые характеристики статистического распределения

7.1. Предмет математической статистики

Математическая статистика — это наука, которая методами теории вероятностей на основании результатов наблюдений изучает закономерности в массовых случайных явлениях. Математическая статистика (не путать со статистикой — разделом экономической теории) указывает способы сбора и группировки статистических данных (результатов наблюдений), разрабатывает методы их обработки для оценки характеристик распределения, для установления зависимости случайной величины от других, для проверки статистических гипотез о виде распределения или значениях его параметров. Математическая статистика возникла и развивалась параллельно с теорией вероятностей.

7.2. Генеральная совокупность и выборка

Значительная часть математической статистики связана с необходимостью описать большую совокупность объектов. Её называют *генеральной совокупностью*.

Если *генеральная совокупность* слишком многочисленна, или её объекты труднодоступны, или имеются другие причины, не позволяющие изучить все объекты, прибегают к изучению какой-то части объектов. Эта выбранная для полного изучения часть называется *выборкой*.

Необходимо, чтобы выборка наилучшим образом представляла генеральную совокупность, т.е. была *репрезентативной* (представительной).

Если генеральная совокупность мала или совсем неизвестна, не удаётся предложить ничего лучшего, чем чисто случайный выбор.

Пример 7.1. *Пусть необходимо оценить качество изделий, выпускаемых определённым цехом машиностроительного предприятия.*

Беляков Н.А. МИРЭА, кафедра высшая математика 2

Для этого выбирают партию изделий и подвергают их контролю с целью дефектирования. Доля бракованных изделий для выбранной партии распространяется затем на всю продукцию цеха.

Здесь генеральная совокупность — все изделия, выпускаемые цехом, выборка — отобранные для проверки изделия.

Пример 7.2. *Пусть необходимо оценить будущий урожай пшеницы. Для этого выбирают небольшой участок поля, например один квадратный метр, и подсчитывают число зерен во всех колосках и их массу. Приближённо весь урожай равен площади поля в метрах, умноженной на массу зерен, собранную с данного участка. Здесь генеральная совокупность — весь ожидаемый урожай, а выборка — урожай, собранный с одного квадратного метра. Если выбрать «плохой» участок (например, близко к краю поля), то оценка урожая будет заниженной. Если же участок имеет преимущества перед другими (например, лучше освещается солнцем), то оценка урожая будет завышенной.*

Пример 7.3. *Производится социологическое исследование с целью прогноза результатов предстоящих выборов мэра города. Здесь генеральная совокупность — все избиратели города, а выборка — число опрошенных респондентов. Большое значение имеет способ, которым получена выборка. Ошибки при выборе способа отбора приводят к тому, что выборка становится нерепрезентативной. Если в качестве респондентов взять, например, сто первых встречных с* 10 *до* 12 *часов дня, то социологи узнают мнение не всех слоев населения, а только домохозяек, направляющихся в это время за покупками.*

Будем проводить испытания и в каждом из них фиксировать значения, которые приняла случайная величина ξ. В результате m испытаний получим выборку n значений, образующих простую статистическую совокупность наблюдений.

Определение 7.1. *Количество наблюдений* n *называется* ***объёмом выборки.***

При большом числе наблюдений (сотни, тысячи) простая статистическая совокупность перестаёт быть удобной формой записи статистического материала — она становится слишком громадной. Для более экономичной записи наблюдаемые значения группируют.

Берлов Н.А. МИРЭА, кафедра высшая математика 2

Пусть в выборке значение x_1 наблюдалось n_1 раз, x_2 — n_2 раз, ..., x_k — n_k раз и $\sum_{i=1}^{k} n_i = n$ — объём выборки.

Определение 7.2. *Наблюдаемые значения x_i называют* ***вариантами****, а их последовательность, записанную в возрастающем порядке —* ***вариационным рядом****. Числа наблюдений $n_1, n_2, \ldots, n_k$ называют* ***частотами****.*

Разность $max(x_i) - min(x_i)$ называется ***размахом вариационного ряда****.*

Статистическим распределением выборки называют перечень вариант и соответствующих им частот — табл. 7.1.

Таблица 7.1

Статистическое распределение			
варианты x_i	x_1	...	x_k
частоты n_i	n_1	...	n_k

7.3. Эмпирическая функция распределения и гистограмма

С каждым испытанием, в котором наблюдается некоторая случайная величина ξ, можно связать случайное событие $\xi = x_i$, но иногда удобнее рассматривать событие $\xi < x_i$.

Определение 7.3. ***Эмпирической (статистической) функцией распределения*** *случайной величины ξ называется функция $F^*(x)$, которая при каждом x равна относительной частоте события $\xi < x$, т.е. отношению n_x — числа наблюдений меньших x к объёму выборки n:*

$$F^*(x) = P^*(\xi < x) = \frac{n_x}{n}\ .$$

Пример 7.4. *Построить эмпирическую функцию распределения для данной выборки:*

Варианты x_i	*1*	*4*	*6*	*7*	*8*	*10*
Частоты n_i	*5*	*10*	*15*	*5*	*10*	*5*

►Объём выборки n равен 5+10+15+5+10+5=50. Наименьшая варианта равна 1, следовательно $F^*(x) = 0$ при $x \leqslant 1$. Значение $x < 3$,

Берков Н.А. МИРЭА, кафедра высшая математика 2

а именно $x = 1$ наблюдалось 5 раз, следовательно $F^*(x) = \frac{5}{50} = 0{,}1$ при $1 < x \leqslant 4$. Значения $x < 6$, а именно $x = 1$ и $x = 4$ наблюдались 5+10=15 раз, следовательно $F^*(x) = \frac{15}{50} = 0{,}3$ при $4 < x \leqslant 6$. Аналогично получаем $F^*(x) = \frac{30}{50} = 0{,}6$ при $6 < x \leqslant 7$ и т.д. Так как 10 — наибольшая варианта, $F^*(x) = 1$ при $x > 10$.

$$F^*(x) = \begin{cases} 0 & \text{при} \quad x \leqslant 1, \\ 0{,}1 & \text{при} \quad 1 < x \leqslant 4, \\ 0{,}3 & \text{при} \quad 4 < x \leqslant 6, \\ 0{,}6 & \text{при} \quad 6 < x \leqslant 7, \\ 0{,}7 & \text{при} \quad 7 < x \leqslant 8, \\ 0{,}9 & \text{при} \quad 8 < x \leqslant 10, \\ 1 & \text{при} \quad x > 10. \end{cases}$$

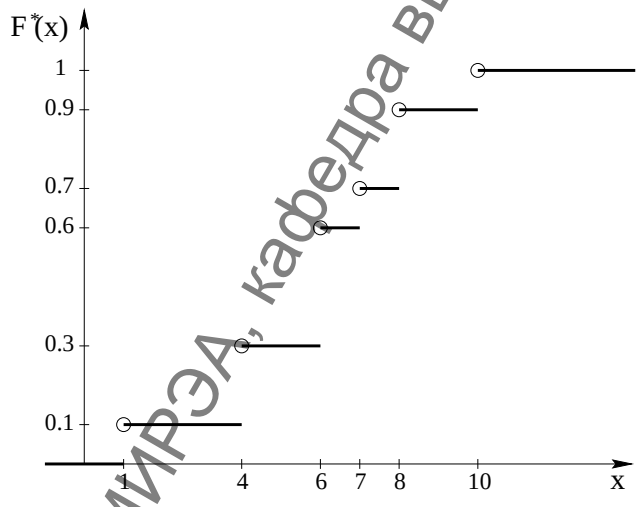

Рис. 25. *Эмпирическая функция распределения*

График найденной функции представлен на рис. 25.◀

Из определения $F^*(x)$ вытекают её свойства:

1) $0 \leqslant F^*(x) \leqslant 1$;
2) $F^*(x)$ — ступенчатая неубывающая функция;
3) если x_1 — наименьшая, а x_k — наибольшая варианты, то $F^*(x) = 0$ при $x \leqslant x_1$ и $F^*(x) = 1$ при $x > x_k$.

Беркович А. МИРЭА, кафедра высшей математики 2

Гистограмма представляет выборку более наглядно. Для построения гистограммы разделим весь диапазон наблюдений на s интервалов вида $(a_{j-1}; a_j]$ и определим количество наблюдений m_j, попавших в j-й интервал. Относительная частота наблюдений, попавших в j-й интервал равна $P_j^* = \dfrac{m_j}{n}$ $(m_1 + \ldots + m_s = n)$, сумма всех частот, очевидно, равна единице. Для построения гистограммы по оси ординат откладываются значения $\dfrac{P_j^*}{\Delta a_j} = \dfrac{m_j}{n \cdot (a_j - a_{j-1})}$. Полученная фигура, состоящая из прямоугольников, называется гистограммой относительных частот. Площадь каждого прямоугольника равна относительной частоте наблюдений, попавших в данный интервал. Для данных примера 7.4 получаются следующие значения:

N п/п	a_{j-1}	a_j	m_j	$P_j^* = \dfrac{m_j}{n}$	$\dfrac{P_j^*}{\Delta a_j}$
1	0	3	5	0.1	1/30
2	3	6	25	0.5	5/30
3	6	9	15	0.3	3/30
4	9	12	5	0.1	1/30

Получившаяся гистограмма представлена на рис. 26.

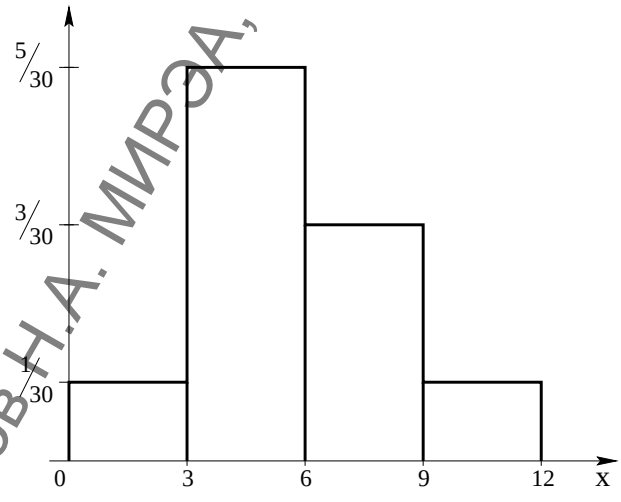

Рис. 26. *Гистограмма относительных частот*

Берков Н.А. МИРЭА, кафедра Высшая математика 2

Другим наглядным способом представления распределения является *полигон относительных частот*. Для его построения по оси абсцисс откладываются варианты, а по оси ординат — относительные частоты (рис. 27), и полученные точки соединяются ломаной линией.

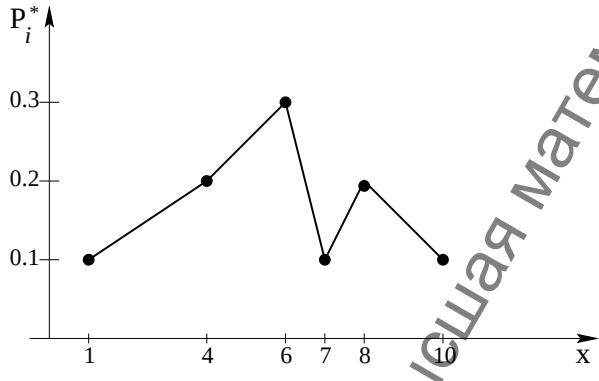

Рис. 27. *Полигон относительных частот*

Для выборки из генеральной совокупности значений непрерывной случайной величины гистограмма является статистическим аналогом плотности распределения, а для дискретной случайной величины полигон относительных частот является статистическим аналогом многоугольника вероятностей. При увеличении объёма выборки эти статистические характеристики в определённом смысле приближаются к своим теоретическим аналогам.

Замечание 7.1. *Наряду с гистограммой и полигоном относительных частот иногда рассматривают соответственно гистограмму и полигон частот, отличающиеся масштабом по оси ординат — все значения по оси ординат умножаются на n — обём выборки. Понятно, что форму получаемых фигур это не изменяет.*

7.4. Числовые характеристики статистического распределения

Статистическая функция распределения и гистограмма являются полными характеристиками результатов наблюдения случайной величины в данной серии испытаний. Однако иногда целесообразно ограничиться более простой, хотя и неполной характеристикой распределения.

Берлов Н.А. МИРЭА, кафедра высшая математика 2

Простейшей характеристикой распределения является *выборочное среднее*, которое для простой статистической совокупности вычисляется по формуле:

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i. \tag{7.1}$$

Если данные сгруппированы, то:

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{k} n_i \cdot x_i. \tag{7.2}$$

Иными словами, выборочное среднее представляет собой среднее взвешенное значение, причём веса равны соответствующим частотам.

Для характеристики разброса значений случайной величины относительно её среднего значения используется *выборочная дисперсия*

$$S^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2 = \overline{(x - \bar{x})^2} \tag{7.3}$$

для простой совокупности и

$$S^2 = \frac{1}{n}\sum_{i=1}^{k} n_i(x_i - \bar{x})^2 \tag{7.4}$$

для сгруппированного распределения.

Очевидно, выборочная дисперсия имеет ту же размерность, что и квадрат случайной величины. Практически удобно пользоваться величиной, имеющей ту же размерность, что и данная случайная величина.

Для этого достаточно из дисперсии извлечь квадратный корень.

Эта величина

$$S = \sqrt{S^2} \tag{7.5}$$

называется *выборочным средним квадратическим отклонением* (СКО).

На практике вместо формулы (7.3) бывает удобнее применять другую:

$$S^2 = \frac{1}{n}\sum_{i=1}^{n} x_i^2 - (\bar{x})^2 = \overline{x^2} - \overline{x}^2 \tag{7.6}$$

для простой совокупности и

$$S^2 = \frac{1}{n}\sum_{i=1}^{k} n_i x_i^2 - (\bar{x})^2 \tag{7.7}$$

Берков Н.А. МИРЭА кафедра высшая математика 2

для сгруппированного распределения.

Докажем формулу (7.6):

$$S^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i^2 - 2x_i\bar{x} + \bar{x}^2) =$$

$$= \frac{1}{n}\sum_{i=1}^{n} x_i^2 - 2\bar{x}\frac{1}{n}\sum_{i=1}^{n} x_i + \bar{x}^2 = \overline{x^2} - 2\bar{x}\cdot\bar{x} + \bar{x}^2 = \overline{x^2} - \overline{x}^2.$$

Модой статистического распределения (обозначается M_O) называется значение которое наиболее часто встречается в исследуемой выборке.

Например, для ряда $2, 2, 4, 5, 5, 5, 5, 6, 6, 7, M_O = 5$. Встречаются распределения, называемые мультимодальными или полимодальными, в которых имеются несколько значений M_0. Например: $1, 2, 2, 3, 4, //5, 5, 6, 6, 7$, модами будут три значения $M_O = \{2, 5, 6\}$.

Медианой (M_e) называется значение, которое разбивает выборку на две равные части. Половина наблюдений лежит ниже медианы, и половина наблюдений лежит выше медианы.

Медиана вычисляется следующим образом. Изучаемая выборка упорядочивается в порядке возрастания. Получаемая последовательность в виде неубывающей последовательности (вариационного ряда) x_i, где $i = 1, ..., n$, где n – объем выборки. Если объем выборки нечетное число, то $M_e = x_{(n+1)/2}$, иначе $M_e = (x_{n/2} + x_{n/2+1})/2$.

Например, для вариационного ряда $\{1, 3, 5, 7, 9, 9, 12\}$ медиана равна третьему элементу $M_e = 7$, а для вариационного ряда $\{1, 3, 5, 7, 9, 12\}$ медиана равна среднему значению второго и третьего элементов $M_e = (5 + 7)/2 = 6$.

Для выборки представленной в виде сгруппированного распределения значений моды и медианы аппроксимируются по следующим формулам

$$M_O = X_O + h\frac{f_{m0} - f_{m0-1}}{2f_{m0} - f_{m0-1} - f_{m0+1}}, \tag{7.8}$$

где X_O – нижнее значение модального интервала; $f_{m0}, f_{m0-1}, f_{m0+1}$ – значение частот в модальном интервале, предыдущем и в следующем интервалах, соответственно; h – размах интервала.

Берков Н.А. МИРЭА, кафедра Высшая математика 2

$$M_e = X_O + h\frac{0{,}5\sum\limits_{k=1}^{s} f_k - \sum\limits_{k=1}^{me-1} f_k}{f_{me}}, \tag{7.9}$$

где X_O – нижняя граница интервала, в котором находится медиана; me – номер медианного интервала; f_{me} – значение частоты в медианном интервале.

Пример 7.5. *Покажем, как построить гистограмму и эмпирическую функцию распределения, вычислить числовые характеристики статистического распределения с помощью программы на Maxima. Сформируем массив x выборки $n = 500$ с помощью датчика случайных чисел.*

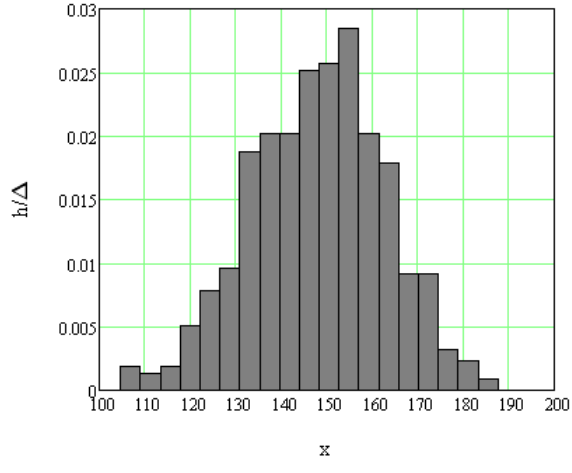

Рис. 28. *Гистограмма относительных частот для примера 7.4*

Maxima-программа:

```
(%i0) kill(all)$   fpprintprec:4$   numer:true$  n:500$
(%i4) load(distrib)$
```

/ Генерируем выборку объёма n псевдослучайными числами. */*

```
(%i5) x:random_normal(145, 15, n)$
```

Веркеев Г.А. Высшая математика 2

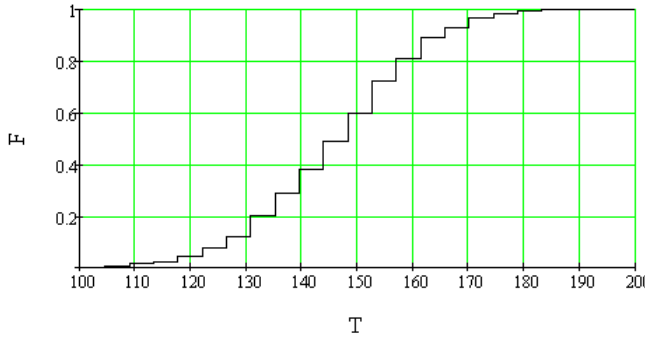

Рис. 29. *Эмпирическая функция распределения для примера 7.5*

*/*Изменяем значение выборки добавлением случайных чисел в диапазоне от 0 до 10.*/*

```
(%i6)x:makelist(x[i]+random(10),i,1,n)$
```

*/*Загружаем библиотеку descriptive.*/*

```
(%i7) load (descriptive)$
```

*/*Строим гистограмму частот.*/*

```
(%i8) histogram(x, nclasses=20, title="закон распределения",
       xlabel="x", ylabel="частоты",
        fill_color=black, fill_density=0.05);
```

*/*Сортируем список в порядке возрастания значений.*/*

```
(%i9) x:sort(x)$
```

*/*Разбиваем выборку обёма n на s интервалов длиной delta. */*

```
(%i10) s:20$  delta:(x[n]-x[1])/s;
```

*/*T - массив узловых координат разбиения. */*

Беркович Л.А. МИРЭА, кафедра ВТ

```
(%i12) T:makelist(x[1]+delta*k,k,-1,s+1);
```

*/*Координаты средних точек отрезков.*/*

```
(%i13) t:makelist((T[m]+T[m+1])/2,m,1,s+2);
```

/ Частоты попадания в соответствующие отрезки.*/*

```
(%i14) h:makelist(0, i, 1, s+2); for j:1 while j<=n do(
      k:fix((x[j] -x[1])/delta)+1,h[k]:h[k]+1);
```

*/*Контрольная сумма объёма выборки.*/*

```
(%i15)  sum(h[i],i,1,s+2);
(%o15) 500
```

*/*Эмпирическая функция распределения.*/*

```
(%i16) F[1]:h[1]; for j:2 while j<=s+2 do(F[j]:F[j-1]+h[j]);
(%i17) listarray(F);
```

*/*График эмпирическая функция распределения.*/*

```
(%i18) wxplot2d([['discrete,makelist([t[j],F[j]/n],j,1,s+2)]],
       [style,[lines,3,5]],  [gnuplot_preamble,"set grid"],
       [ylabel,""])$
```

/ Найдем также некоторые числовые характеристики данной выборки: Выборочное среднее.*/*

```
(%i19) mean(x);
```

/ Выборочная дисперсия.*

```
(%i20) var(x);
```

/ Выборочное среднее квадратическое отклонение.*/*

```
(%i21) std(x);
```

/ Несмещенная выборочная дисперсия.*/*

```
(%i22) var1(x);
```

/ Несмещенное среднее квадратическое отклонение.*/*

```
(%i23) std1(x);
```

/ Медиана.*/*

Беркович А. МИРЭА, кафедра высшая математика 2

```
(%i24) median(x);
```

7.5. Точечные оценки параметров распределения

Выборочное среднее, выборочная дисперсия и СКО являются примерами точечных оценок параметров распределения.

Определение 7.4. *Точечной оценкой* $\widetilde{a}_n$ *неизвестного параметра* a *распределения случайной величины* ξ *называется функция от наблюдений:*

$$\widetilde{a}_n = \widetilde{a}(x_1, \ \ldots, x_n).$$

Для изучения свойств этой оценки её рассматривают как функцию от n независимых случайных величин $\xi_1, \ldots, \xi_n$, имеющих такое же распределение, что и ξ; $x_1, \ldots, x_n$ в этом случае рассматриваются как наблюдения над этими случайными величинами: x_1 — полученное значение ξ_1, x_2 — наблюденное значение ξ_2 и т.д. Сама оценка $\widetilde{a}_n$ в этом случае является случайной величиной.

Перечислим свойства точечной оценки $\widetilde{a}_n$, которые могут считаться «хорошими».

Определение 7.5. *Оценка* $\widetilde{a}_n$ *называется* ***состоятельной****, если при* $n \to \infty$ *она сходится по вероятности к оцениваемому параметру* a*:*

$$\lim_{n\to\infty} P\{|\widetilde{a}_n - a| < \varepsilon\} = 1 \ \textit{для} \ \forall\ \varepsilon > 0.$$

Определение 7.6. *Оценка* $\widetilde{a}_n$ *называется* ***несмещенной****, если её математическое ожидание равно оцениваемому параметру* a*:*

$$M(\widetilde{a}_n) = a.$$

Иногда точечные оценки обладают более слабым свойством: их смещение $M(\widetilde{a}_n) - a$ стремится к нулю при $n \to \infty$. Такие оценки называются асимптотически несмещёнными.

Определение 7.7. *Несмещённая оценка* $\widetilde{a}_n$ *называется* ***эффективной****, если её дисперсия наименьшая по сравнению с другими несмещёнными оценками.*

На практике оценка не всегда удовлетворяет всем этим требованиям одновременно.

Беркович Н.А., МИРЭА, кафедра высшая математика 2

Пример 7.6. *Доказать, что выборочное среднее $\bar{x}$ является несмещённой и состоятельной оценкой для математического ожидания (генерального среднего) случайной величины.*

Р е ш е н и е: Обозначим $M(\xi) = a,\ D(\xi) = \sigma^2$. Рассматривая $\bar{x}$ как случайную величину, найдем её математическое ожидание. При этом, как было отмечено ранее, считаем
$M(\xi_1) = \ldots = M(\xi_n) = a, \quad D(\xi_1) = \ldots = D(\xi_n) = \sigma^2$.

$$M(\xi) = M\left(\frac{\sum\limits_{i=1}^{n}\xi_i}{n}\right) = \frac{\sum\limits_{i=1}^{n}M(\xi_i)}{n} = \frac{na}{n} = a.$$

Несмещённость выборочного среднего доказана. Оценим теперь дисперсию выборочного среднего:

$$D(\xi) = D\left(\frac{\sum\limits_{i=1}^{n}\xi_i}{n}\right) = \frac{\sum\limits_{i=1}^{n}D(\xi_i)}{n^2} = \frac{n\sigma^2}{n^2} = \frac{\sigma^2}{n}.$$

В соответствии с неравенством Чебышева (теорема 6.1) получаем $\forall\ \varepsilon > 0$:

$$1 \geqslant P\{|\xi - M(\xi)| < \varepsilon\} \geqslant 1 - \frac{\sigma^2/n}{\varepsilon^2}.$$

Заменяя $M(\xi) = a$ и переходя к пределу при $n \to \infty$, получаем

$$1 \geqslant \lim_{n\to\infty} P\{|\xi - a| < \varepsilon\} \geqslant 1,$$

откуда получаем:

$$\lim_{n\to\infty} P\{|\xi - a| < \varepsilon\} = 1.$$

Это равенство и означает состоятельность оценки $\bar{x}$.

Замечание 7.2. *Можно доказать, что выборочное среднее будет эффективной оценкой математического ожидания в случае, когда случайная величина имеет нормальное распределение.*

Аналогично доказывается, что выборочная дисперсия S^2 является состоятельной и смещённой оценкой дисперсии σ^2:

$$M(S^2) = \frac{n-1}{n}\,\sigma^2. \tag{7.10}$$

Примем это без доказательства.

Берков Н.А. МИРЭА, кафедра Высшая математика 2

При малых объёмах выборки n для оценки дисперсии σ^2 используют исправленную выборочную дисперсию S^{*^2}:

$$S^{*^2} = \frac{n}{n-1} S^2 = \frac{1}{n-1} \sum_{i=1}^{n} (x_i - \bar{x})^2. \qquad (7.11)$$

Оценка S^{*^2} является несмещённой, состоятельной оценкой дисперсии σ^2.

Формула (7.11) позволяет вычислять S^{*^2} для простой совокупности. Для сгруппированных данных используют аналогичную формулу (7.12):

$$S^{*^2} = \frac{1}{n-1} \sum_{i=1}^{k} n_i (x_i - \bar{x})^2. \qquad (7.12)$$

Замечание 7.3. *Исправленное СКО S^* является смещённой оценкой СКО S.*

7.6. Распределения, используемые в статистике

Познакомимся с некоторыми непрерывными распределениями, которые применяются в математической статистике.

Распределение χ^2 (хи-квадрат).

Пусть имеется n независимых стандартных нормальных случайных величин $\xi_1, \xi_2, \ldots, \xi_n, \quad \xi_i \sim N(0;1),\ i = 1, \ldots, n$.

Определение 7.8. *Распределение случайной величины $\chi^2_n = \sum_{i=1}^{n} \xi_i^2$ называется χ^2—распределением с n степенями свободы.*

Очевидно, что случайная величина $\chi^2_n \geqslant 0$.

Плотность этого распределения имеет вид:

$$f(x) = \begin{cases} \dfrac{x^{\frac{k}{2}-1}}{\Gamma\left(\frac{k}{2}\right) 2^{\frac{k}{2}}} e^{-\frac{x}{2}} & \text{при} \quad x \geqslant 0, \\ 0 & \text{при} \quad x < 0. \end{cases}$$

Здесь $\Gamma(x) = \int\limits_0^\infty t^{x-1} \cdot e^{-t} dt$ — гамма функция, являющаяся обобщением понятия факториала: $\Gamma(x) = (x-1)!$ при $x \geqslant 1$.

Берков Н.А. МИРЭА, кафедра высшая математика 2

Замечание 7.4. *Если случайные величины $\xi_1, \ldots, \xi_n$ связаны какой-нибудь зависимостью, например $\xi_1 + \ldots + \xi_n = n \cdot \overline{x}$, то число степеней свободы уменьшается, случайная величина $\sum_{i=1}^{n} \xi_i^2$ будет иметь распределение χ^2_{n-1}.*

Распределение Стьюдента.

Пусть имеется $n+1$ независимая стандартная случайная величина $\zeta, \xi_1, \ldots, \xi_n$.

Определение 7.9. *Распределение случайной величины*

$$t = \frac{\zeta}{\sqrt{\dfrac{\chi_n^2}{n}}}$$

называется распределением Стьюдента с n степенями свободы.

Плотность этого распределения имеет вид:

$$f(x) = \frac{\left(1 + \frac{x^2}{2}\right)^{\frac{n+1}{2}}}{2^{\frac{n}{2}} \Gamma\left(\frac{n}{2}\right) \sqrt{\pi n}}.$$

Поскольку распределение симметрично относительно нуля (плотность — чётная функция), математическое ожидание равно нулю.

Стьюдент — псевдоним английского статистика Госсета.

F—Распределение Фишера--Снедекора.

Пусть имеется $n + k$ независимых стандартных величин: $\xi_1, \ldots, \xi_n$; $\zeta_1, \ldots, \zeta_k$; $\xi_i \sim N(0;1),\ i = 1, \ldots, n$; $\zeta_i \sim N(0;1),\ j = 1, \ldots, k$.

Определение 7.10. *Распределение случайной величины*

$$F_{n,k} = \frac{\dfrac{\chi_n^2}{n}}{\dfrac{\chi_k^2}{k}}$$

называется F—распределением Фишера—Снедекора (распределением Фишера или F—распределением) с n, k степенями свободы.

Очевидно, что случайная величина $F_{n,k} \geqslant 0$.

Берков Н.А. МИРЭА, кафедра высшая математика 2

Плотность этого распределения имеет вид:

$$f(x) = \begin{cases} \dfrac{\Gamma\left(\dfrac{m+k}{2}\right)}{\Gamma\left(\dfrac{n}{2}\right)\cdot\Gamma\left(\dfrac{k}{2}\right)}\left(\dfrac{n}{k}\right)^{\frac{n}{2}}\cdot x^{\frac{n}{2}-1}\left(1+\dfrac{n}{k}x\right)^{-\frac{n+k}{2}} & \text{при } x \geqslant 0, \\ 0 & \text{при } x < 0. \end{cases}$$

Для всех этих распределений имеются таблицы плотности и функции распределения; их можно также вычислить с помощью прикладных программ на ЭВМ (таких, как Excel, Mathcad, Maxima и проч.).

7.7. Интервальные оценки параметров распределения

Наряду с рассмотренными точечными оценками, определяемыми одним числом, используют интервальные оценки неизвестных параметров, определяемые двумя числами — концами интервала, дающими вероятностную оценку сверху и снизу неизвестного параметра распределения.

Интервальные оценки целесообразно применять при малом объёме выборки, когда дисперсия точечной оценки велика и она может сильно отличаться от оцениваемого параметра.

Определение 7.11. *Доверительным интервалом для несмещённого параметра a называют интервал $(a_1; a_2)$ со случайными границами, зависящими от наблюдений: $a_1 = a_1(x_1, \ldots, x_n)$, $a_2 = a_2(x_1, \ldots, x_n)$, накрывающий неизвестный параметр с заданной вероятностью γ: $P\{a \in (a_1; a_2)\} = \gamma$. Вероятность γ называется доверительной вероятностью или надежностью доверительного интервала.*

Обычно γ задают равным 0,95; 0,99 и более.

Доверительный интервал для неизвестного математического ожидания нормального распределения при известной дисперсии имеет вид:

$$\left(\bar{x} - \tau_{\gamma/2}\frac{\sigma}{\sqrt{n}};\ \bar{x} + \tau_{\gamma/2}\frac{\sigma}{\sqrt{n}}\right), \tag{7.13}$$

где величина $\tau_{\gamma/2}$ определяется из уравнения:

$$\Phi(\tau_{\gamma/2}) = \frac{\gamma}{2} \tag{7.14}$$

по таблицам функции Лапласа или с помощью компьютера, а $\bar{x}$ — выборочное среднее.

Беркович Н.А. МИРЭА, кафедра высшая математика 2

Замечание 7.5. *При возрастании объёма выборки n, как видно из (7.13), доверительный интервал уменьшается. При увеличении надежности γ увеличивается величина $\tau_{\gamma/}$, т.к. функция Лапласа в (7.14) возрастающая; следовательно, увеличивается и доверительный интервал (7.13).*

Для получения доверительного интервала (7.13) заметим, что если независимые случайные величины $\xi_i \sim N(a;\sigma)$, $i = 1, \ldots, n$, то среднее арифметическое $\bar{\xi} = (\xi_1 + \ldots \xi_n)/n$ тоже распределено нормально с параметрами:

$$M(\bar{\xi}) = a, \quad \sigma(\bar{\xi}) = \frac{\sigma}{\sqrt{n}}. \tag{7.15}$$

Формулы (7.15) были получены в примере 7.1. Будем искать доверительный интервал для a в виде:

$$P\{|\bar{\xi} - a| < \varepsilon\} = \gamma, \tag{7.16}$$

где γ — заданная доверительная вероятность. Для определения ε воспользуемся формулой (5.20), которая в данном случае с учётом (7.15) принимает вид:

$$P\{|\bar{\xi} - a| < \varepsilon\} = 2\Phi\Big(\frac{\varepsilon}{\sigma/\sqrt{n}}\Big).$$

Найдём ε из уравнения:

$$2\Phi\Big(\frac{\varepsilon}{\sigma/\sqrt{n}}\Big) = \gamma \implies \Phi\Big(\frac{\varepsilon}{\sigma/\sqrt{n}}\Big) = \frac{\gamma}{2} \implies \frac{\varepsilon}{\sigma/\sqrt{n}} = \tau_{\frac{\gamma}{2}} \implies$$
$$\varepsilon = \tau_{\frac{\gamma}{2}} \cdot \frac{\sigma}{\sqrt{n}}.$$

С учётом полученной величины ε доверительный интервал (7.16) принимает вид (7.13).

Пример 7.7. *Найти доверительный интервал для неизвестного математического ожидания a нормально распределённой случайной величины со средним квадратическим отклонением $\sigma = 2$ по выборке объёма $n = 64$ с выборочным средним $\bar{x} = 5{,}2$. Надежность доверительного интервала $\gamma = 0{,}95$.*

►Из уравнения (7.14) по таблице приложения 2 находим для $\frac{\gamma}{2} = 0{,}475$ $\tau_{\frac{\gamma}{2}} = 1{,}96$. Подставляя найденное значение в (7.13), получаем (4,71;5,69).◄

Ответ: (4,71;5,69).

Беляков Н.А. МИРЭА, кафедра высшая математика 2

Доверительный интервал для неизвестного математического ожидания нормального распределения при неизвестной дисперсии имеет вид:

$$\left(\bar{x} - t_\gamma \frac{S^*}{\sqrt{n}};\ \bar{x} + t_\gamma \frac{S^*}{\sqrt{n}}\right), \tag{7.17}$$

где величина t_γ определяется по таблице приложения 3 критических точек распределения Стьюдента для $\alpha = 1 - \gamma$ и $k = n - 1$ или с помощью компьютера из уравнения для функции распределения Стьюдента $F_{st}(x)$ с $n - 1$ степенью свободы:

$$F_{st}(t_\gamma) = \frac{1 + \gamma}{2}, \tag{7.18}$$

где $\bar{x}$ и S^* — соответственно выборочное среднее и исправленное СКО.

Для получения доверительного интервала (7.17) примем без доказательства, что если независимые случайные величины $\xi_i \sim N(a; \sigma)$, $i = 1,\ \ldots, n$, то случайная величина

$$t = \frac{\bar{\xi} - a}{S^*/\sqrt{n}} \tag{7.19}$$

имеет распределение Стьюдента с $n - 1$ степенью свободы (см п. 93.4).

Обозначим t_γ значение, при котором с вероятностью γ выполняется следующее неравенство:

$$P\{|t| < t_\gamma\} = \gamma. \tag{7.20}$$

С учётом четности плотности распределения Стьюдента $f_{st}(t)$ значение t_γ определяется из условия:

$$P\{|t| < t_\gamma\} = \gamma \iff P\{|t| > t_\gamma\} = 1 - \gamma \implies P\{t > t_\gamma\} = \frac{1 - \gamma}{2} \iff$$

$$\iff 1 - F_{st}(t_\gamma) = \frac{1 - \gamma}{2} \iff F_{st}(t_\gamma) = \frac{1 + \gamma}{2}.$$

Подставляя в (7.20) выражение (7.19), получаем:

$$P\left\{\left|\frac{\bar{\xi} - a}{S^*/\sqrt{n}}\right| < t_\gamma\right\} = \gamma \iff P\left\{-t_\gamma < \frac{\bar{\xi} - a}{S^*/\sqrt{n}} < t_\gamma\right\} = \gamma,$$

откуда получаем для a доверительный интервал в виде (7.17).

Замечание 7.6. *В некоторых пакетах прикладных программ для ЭВМ, например в Excel, под распределением Стьюдента понимается* $1 - F_{st}(x)$. *Поэтому, задавая значение* $1 - \gamma$ *и число свободы, с помощью обратной функции можно сразу получить значение* t_γ *для*

Бердов Н.А. МИРЭА кафедра Высшая математика 2

двустороннего интервала (без использования (7.18)). Указанные особенности можно узнать из инструкций к программам.

Пример 7.8. *Найти доверительный интервал для неизвестного математического ожидания a нормально распределённой случайной величины с выборочным средним $\bar{x} = 10{,}5$ и исправленным СКО $S^* = 1{,}6$ по выборке объёма $n = 16$. Надежность доверительного интервала $\gamma = 0{,}99$.*

►По таблице приложения 3 для числа степеней свободы $k = n-1 = 15$ и $\alpha = 1 - \gamma = 0{,}01$ находим $t_\gamma = 2{,}95$. Подставляя полученное значение в (7.17), получаем значение для радиуса доверительного интервала ε:

$$\varepsilon = t_\gamma \frac{S^*}{\sqrt{n}} = 2{,}95 \frac{1{,}6}{\sqrt{16}} = 2{,}95 \cdot 0{,}4 = 1{,}18.$$

Находим доверительный интервал $(10{,}5-1{,}18;\, 10{,}5+1{,}18) = (9{,}32;\, 11{,}68)$.◄

Ответ: (9,32; 11,68).

Берков Н.А. МИРЭА, кафедра высшая математика 2

7.8. Выборочный коэффициент корреляции

Рассмотрим выборку объёма n из генеральной совокупности значений двумерной случайной величины $(\xi;\zeta)$, т.е. n пар наблюдений $(x_i;y_i)$. Поскольку многие значения в этой выборке могут повторяться, их заносят в так называемую корреляционную таблицу (табл. 7.4). В первом столбце этой таблицы перечислены значения x_i, во втором — y_i в виде вариационных рядов. На пересечении i-й строки

Таблица 7.4

Корреляционная таблица					
$\xi\backslash\zeta$	y_1	y_2		y_s	$n_{i\cdot}$
x_1	n_{11}	n_{12}		n_{1s}	$n_{1\cdot}$
x_2	n_{21}	n_{22}		n_{2s}	$n_{2\cdot}$
x_k	n_{k1}	n_{k2}		n_{ks}	$n_{k\cdot}$
$n_{\cdot j}$	$n_{\cdot 1}$	$n_{\cdot 2}$		$n_{\cdot s}$	n

и j-го столбца — соответствующая частота n_{ij}, т.е. количество раз, которое наблюдение $(x_i;y_j)$ встретилось в выборке. При обработке корреляционной таблицы в последнем столбце указывают сумму частот по строкам $n_{i\cdot} = \sum\limits_{j=1}^{s} n_{ij}$, а в последней строке — сумму частот по столбцам $n_{\cdot j} = \sum\limits_{i=1}^{k} n_{ij}$. Сумма всех элементов последнего столбца или строки даст объём выборки

$$n = \sum_{i=1}^{k}\sum_{j=1}^{s} n_{ij} = \sum_{i=1}^{k} n_{i\cdot} = \sum_{j=1}^{s} n_{\cdot j}.$$

Первый и последний столбцы корреляционной таблицы образуют статистическое распределение выборки случайной величины ξ, а первая и последняя строки образуют выборку случайной величины ζ. Обработав их, как описано в п. 96.4 предыдущей лекции, получим числовые характеристики

$$\overline{x} = \frac{\sum\limits_{i=1}^{k} n_{i\cdot}x_i}{n}, \quad \overline{x^2} = \frac{\sum\limits_{i=1}^{k} n_{i\cdot}x_i^2}{n}, \quad S_x^2 = \overline{x^2} - \overline{x}^2,$$

Берков Н.А. МИРЭА, кафедра высшая математика 2

$$\overline{y} = \frac{\sum\limits_{j=1}^{s} n_{\cdot j} y_j}{n}, \quad \overline{y^2} = \frac{\sum\limits_{j=1}^{s} n_{\cdot j} y_j^2}{n}, \quad S_y^2 = \overline{y^2} - \overline{y}^2.$$

Определение 7.12. *Выборочным коэффициентом корреляции* r^*_{xy} *называется:*

$$r^*_{xy} = \frac{\overline{xy} - \overline{x} \cdot \overline{y}}{S_x \cdot S_y}, \quad \textit{где} \tag{7.21}$$

$$\overline{xy} = \frac{\sum\limits_{i=1}^{k} \sum\limits_{j=1}^{s} n_{ij} x_i y_j}{n}. \tag{7.22}$$

Выборочный коэффициент корреляции является статистической оценкой коэффициента корреляции, рассмотренного в лекции 95, и он обладает следующими свойствами, которые мы приведем без доказательства:

1) $r^*_{xy} = r^*_{yx}$;

2) Выборочный коэффициент корреляции находится в пределах от -1 до 1: $-1 \leqslant r^*_{xy} \leqslant 1$;

3) $|r^*_{xy}| = 1$ тогда и только тогда, когда между значениями x_i и y_i имеется линейная зависимость. Чем ближе r^*_{xy} к нулю, тем хуже эта зависимость аппроксимируется линейной.

Определение 7.13. *Условным средним* $\overline{y_x}$ *называют среднее арифметическое значений* ζ *при фиксированном значении* $\xi = x$.

Для корреляционной таблицы 7.4 условное среднее $\overline{y_x}$ *получается усреднением значений* ζ *по строке, соответствующей* $\xi = x$. *Так, например,* $\overline{y_{x_1}} = \dfrac{\sum\limits_{j=1}^{S} y_j n_{1j}}{n_{1\cdot}}$. *Аналогично определяется условное среднее* $\overline{x_y}$.

Ранее было введено понятие регрессии ζ на ξ: $M(\zeta/\xi = x) = f_{\zeta/\xi}(x)$ и ξ на ζ: $M(\xi/\zeta = y) = \Psi_{\xi/\zeta}(y)$ и получены формулы (6.23) и (??) для прямых среднеквадратической регрессии. Ниже будут введены их статистические аналоги.

Берков Н.А. МИРЭА, кафедра высшая математика 2

7.9. Проверка статистических гипотез

Основные понятия. Проверка гипотезы о значимости выборочного коэффициента корреляции. Сравнение двух математических ожиданий. Сравнение математического ожидания с заданным значением. Сравнение вероятности с заданным значением

Определение 7.14. *Статистической называется гипотеза о виде распределения или о значениях его параметров.*

Гипотезы будем обозначать $H_0,\ H_1,\ H_2,\ \ldots$

Различают проверяемую или основную гипотезу H_0 и альтернативную или конкурирующую H_1, которая должна противоречить основной.

Пример 7.9. *Проверяемая гипотеза H_0 состоит в том, что математическое ожидание случайной величины ξ равно заданному значению a_0. $H_0:\ M(\xi) = a_0$. Альтернативная $H_1:\ M(\xi) > a_0$.*

Для проверки статистической гипотезы на основании выборки $x_1,\ x_2,\ \ldots,\ x_n$ вычисляют значение критерия, зависящего от наблюдений:

$$T = T(x_1,\ x_2,\ \ldots\ x_n).$$

Всё множество значений критерия делится на так называемую критическую область, при попадании в которую критерия проверяемая гипотеза отвергается, и область принятия гипотезы.

При принятии решения о справедливости гипотезы H_0 возможны следующие ошибки:

- гипотеза H_0 отвергается, хотя на самом деле она верна (ошибка первого рода) ;
- гипотеза H_0 принимается, хотя на самом деле она не верна, а справедлива гипотеза H_1 (ошибка второго рода) .

Наряду с этим возможны следующие правильные решения:

- гипотеза H_0 принимается и она действительно верна;
- гипотеза H_0 отвергается и на самом деле справедлива гипотеза H_1.

Определение 7.15. *Вероятность ошибки первого рода называется уровнем значимости критерия и обычно обозначается α.*

Беркович А. МИРЭА, кафедра Высшая математика 2

Вероятность правильно отвергнуть проверяемую гипотезу называется мощностью критерия и обычно обозначается β, тогда вероятность ошибки второго рода равна $1-\beta$.

Одновременно уменьшить вероятности ошибок первого и второго рода можно только увеличив объём выборки n. При фиксированном n обычно задают допустимый уровень ошибки первого рода α и стараются минимизировать вероятность ошибки второго рода $1-\beta$, т.е. максимизировать мощность критерия β.

На практике при проверке статистической гипотезы на основании наблюдений вычисляют наблюдаемое значение критерия $T_{\text{набл}}$ и по заданному уровню значимости α определяют границы критической области — критические точки.

Если критическая область правосторонняя, т.е. $(t_{\text{кр2}};\ +\infty)$, при выполнении условия $T_{\text{набл}} > t_{\text{кр2}}$ делают вывод: проверяемая гипотеза H_0 отвергается с уровнем значимости α в пользу гипотезы H_1; если это условие не выполняется, т.е. $T_{\text{набл}} \leqslant t_{\text{кр2}}$, делают более осторожный вывод: нет оснований для того, чтобы отвергнуть гипотезу H_0 в пользу гипотезы H_1 с уровнем значимости α.

Если критическая область левосторонняя, т.е. $(-\infty;\ t_{\text{кр1}})$, гипотеза H_0 отвергается при выполнении условия $T_{\text{набл}} < t_{\text{кр1}}$. В случае двусторонней критической области вида $(-\infty;\ t_{\text{кр1}}) \cup (t_{\text{кр2}};\ +\infty)$ гипотеза H_0 отвергается при выполнении условия $T_{\text{набл}} < t_{\text{кр1}}$ или $T_{\text{набл}} > t_{\text{кр2}}$.

7.10. Проверка гипотезы о значимости выборочного коэффициента корреляции

Пусть на основании данных корреляционной таблицы по выборке объёма n независимых наблюдений над нормально распределёнными случайными величинами найден выборочный коэффициент корреляции r^*_{xy}, который оказался отличным от нуля. Так как выборка отобрана случайно, возникает вопрос о том, будет ли отличен от нуля теоретический коэффициент корреляции $r_{\xi\zeta}$, к которому сходится выборочный коэффициент при $n \to \infty$.

Необходимо при заданном уровне значимости α проверить гипотезу $H_0:\ r_{\xi\zeta} = 0$ при альтернативной гипотезе $H_1:\ r_{\xi\zeta} \neq 0$.

Если H_0 отвергается, это означает, что выборочный коэффициент корреляции значимо отличается от нуля, а случайные величины ξ и ζ

Беляев Н.Г. МИРЭА, кафедра высшая математика 2

коррелированы, т.е. в той или иной степени связаны линейной зависимостью. Если H_0 принимается, это означает, что выборочный коэффициент корреляции незначимо отличается от нуля, а случайные величины ξ и ζ некоррелированы, т.е. не связаны линейной зависимостью.

В качестве критерия для проверки H_0 выбирается случайная величина

$$T = r^*_{xy} \frac{\sqrt{n-2}}{\sqrt{1 - {r^*_{xy}}^2}} , \tag{7.23}$$

где r^*_{xy} вычисляется по формуле (7.21). При справедливости гипотезы H_0 величина T имеет так называемое распределение Стьюдента с $n-2$ степенями свободы. Критическая область для рассматриваемой гипотезы H_1 будет двусторонней, $t_{\text{кр1}} = -t_{\text{кр2}}$. Критическая точка $t_{\text{кр2}}$ определяется по заданным уровню значимости α и числу степеней $n-2$ по специальным таблицам (приложение 3) или с помощью обратной к функции распределения Стьюдента, имеющейся, например, среди статистических функций Excel для $\alpha/2$ и $n-2$ степеней свободы. По формуле (7.23) для данных наблюдений определяем значение критерия $T_{\text{набл}}$.

Если $|T_{\text{набл}}| > t_{\text{кр2}}$, гипотеза H_0 отвергается с уровнем значимости α, если $|T_{\text{набл}}| \leqslant t_{\text{кр2}}$ — нет оснований отвергнуть H_0.

Берков Н.А. МИРЭА, кафедра Высшая математика 2

7.11. Сравнение двух математических ожиданий

Пусть имеются две независимые выборки объёмов n и m из нормальных совокупностей с известными дисперсиями σ_1^2 и σ_2^2. Требуется по найденным выборочным средним $\bar{x}$ и $\bar{y}$ с уровнем значимости α проверить нулевую гипотезу H_0 о равенстве теоретических математических ожиданий:

$$H_0: \quad M(\xi) = M(\zeta).$$

Заметим, что в силу несмещённости оценок $\bar{x}$ и $\bar{y}$ следует, что нулевую гипотезу можно записать и так:

$$H_0: \quad M(\bar{\xi}) = M(\bar{\zeta}).$$

Другими словами, требуется проверить значимо или нет отличаются между собой выборочные средние. В качестве критерия проверки гипотезы примем величину:

$$Z = \frac{\bar{x} - \bar{y}}{\sqrt{\dfrac{\sigma_1^2}{n} + \dfrac{\sigma_2^2}{m}}}. \tag{7.24}$$

Для изучения её свойств рассмотрим соответствующую случайную величину:

$$Z = \frac{\bar{\xi} - \bar{\zeta}}{\sqrt{\dfrac{\sigma_1^2}{n} + \dfrac{\sigma_2^2}{m}}}, \text{ где } \quad \bar{\xi} = \frac{\sum\limits_{i=1}^{n} \xi_i}{n},\ \bar{\zeta} = \frac{\sum\limits_{i=1}^{m} \zeta_i}{m}\ .$$

Если верна гипотеза H_0, т.е. $\xi_i \sim N(a;\sigma_1)$, $\zeta_i \sim N(a;\sigma_2)$, то $Z \sim N(0;1)$.

Действительно, Z является линейной комбинацией нормально распределённых случайных величин и поэтому сама распределена нормально. Её математическое ожидание и дисперсия равны:

$$M(Z) = \big(M(\bar{\xi}) - M(\bar{\zeta})\big) \Big/ \sqrt{\frac{\sigma_1^2}{n} + \frac{\sigma_2^2}{m}} =$$

$$= \left(\sum_{i=1}^{n} M(\xi_i)/n - \sum_{i=1}^{m} M(\zeta_i)/m\right) \Big/ \sqrt{\frac{\sigma_1^2}{n} + \frac{\sigma_2^2}{m}} =$$

$$= \left(\frac{na}{n} - \frac{ma}{m}\right) \Big/ \sqrt{\frac{\sigma_1^2}{n} + \frac{\sigma_2^2}{m}} = 0,$$

Берков Н.А. МИРЭА, кафедра высшая математика 2

$$D(Z) = \big(D(\bar{\xi}) + D(\bar{\zeta})\big) \Big/ \left(\frac{\sigma_1^2}{n} + \frac{\sigma_2^2}{m}\right) =$$

$$= \left(\sum_{i=1}^{n} D(\xi_i)/n^2 + \sum_{i=1}^{m} D(\zeta_i)/m^2\right) \Big/ \left(\frac{\sigma_1^2}{n} + \frac{\sigma_2^2}{m}\right) =$$

$$= \left(\frac{n\sigma_1^2}{n^2} + \frac{m\sigma_2^2}{m^2}\right) \Big/ \left(\frac{\sigma_1^2}{n} + \frac{\sigma_2^2}{m}\right) = 1.$$

Поэтому, в зависимости от конкурирующей гипотезы, решающее правило выглядит следующим образом:

- $H_0: \quad M(\xi) = M(\zeta), \qquad H_1: \quad M(\xi) \neq M(\zeta).$

  Критическая область двусторонняя с вероятностью $\alpha/2$ попадания в каждую половину в случае справедливости H_0. Из уравнения $F_{\text{ст}}(Z_{\text{кр}}) = 1 - \alpha/2$, где $F_{\text{ст}}(Z)$ — функция распределения стандартного нормального закона, находим значение $Z_{\text{кр}}$, вычисляем по данным наблюдениям значение критерия $Z_{\text{набл}}$ и если $|Z_{\text{набл}}| > Z_{\text{кр}}$, то отвергаем гипотезу H_0 с уровнем значимости α. Если $|Z_{\text{набл}}| \leqslant Z_{\text{кр}}$, у нас нет оснований отвергнуть гипотезу H_0 в пользу данной гипотезы H_1.

  На практике уравнение $F_{\text{ст}}(Z_{\text{кр}}) = 1 - \alpha/2$ решают или с помощью ЭВМ (например, Excel), или по таблице приложения 2 и уравнения (7.25), т.к.

$$F_{\text{ст}}(Z_{\text{кр}}) = \Phi(Z_{\text{кр}}) + 0{,}5 \implies F_{\text{ст}}(Z_{\text{кр}}) = 1 - \frac{\alpha}{2} \iff$$

$$\iff \Phi(Z_{\text{кр}}) + 0{,}5 = 1 - \frac{\alpha}{2} \iff$$

$$\Phi(Z_{\text{кр}}) = \frac{1}{2} - \frac{\alpha}{2}; \tag{7.25}$$

- $H_0: \quad M(\xi) = M(\zeta), \qquad H_2: \quad M(\xi) > M(\zeta).$

  Критическая область правосторонняя с вероятностью α попадания в неё в случае справедливости H_0. Из уравнения $F_{\text{ст}}(Z_{\text{кр}}) = 1 - \alpha$ находим значение $Z_{\text{кр}}$, вычисляем по формуле (7.24) $Z_{\text{набл}}$ и если $Z_{\text{набл}} > Z_{\text{кр}}$, то отвергаем гипотезу H_0 с уровнем значимости α. Если $Z_{\text{набл}} \leqslant Z_{\text{кр}}$, то нет оснований отвергнуть гипотезу H_0. На практике $Z_{\text{кр}}$ находят или с помощью ЭВМ или по таблице приложения 2, из уравнения

Беркович А. МИРЭА, кафедра высшая математика 2

(7.26) т.к.

$$F_{\text{ст}}(Z_{\text{кр}}) = 1-\alpha \iff \Phi(Z_{\text{кр}}) + 0{,}5 = 1-\alpha \iff$$

$$\Phi(Z_{\text{кр}}) = \frac{1}{2} - \alpha; \tag{7.26}$$

- $H_0: \quad M(\xi) = M(\zeta), \qquad H_3: \quad M(\xi) < M(\zeta).$

Критическая область левосторонняя с вероятностью α попадания в неё в случае справедливости H_0. Из уравнения $F_{\text{ст}}(Z'_{\text{кр}}) = \alpha$ находим значение $Z'_{\text{кр}}$.

В силу симметрии нормального распределения относительно нуля на практике находят значение $Z_{\text{кр}}$ из уравнения (7.26) и берут $Z'_{\text{кр}} = -Z_{\text{кр}}$. Если $Z_{\text{набл}} < -Z_{\text{кр}}$, гипотезу H_0 отвергают с уровнем значимости α, если $Z_{\text{набл}} \geqslant -Z_{\text{кр}}$, то нет оснований отвергнуть H_0.

Замечание 7.7. *Если независимые выборки достаточно большие, указанный критерий можно применять для случая неизвестных дисперсий и не обязательно нормального распределения совокупностей. В этом случае вместо формулы (7.24) используют формулу (7.27) для вычисления критерия Крамера–Уэлча:*

$$Z = \frac{\bar{x} - \bar{y}}{\sqrt{\dfrac{S_1^{*2}}{n} + \dfrac{S_2^{*2}}{m}}}. \tag{7.27}$$

7.12. Сравнение математического ожидания с заданным значением

Пусть имеется выборка объёма n нормальной совокупности с известной дисперсией σ^2. Требуется по найденной выборочной средней с уровнем значимости α проверить гипотезу H_0 о равенстве неизвестного математического ожидания $M(\xi)$ заданному значению a_0:

$$H_0: \quad M(\xi) = a_0.$$

В силу несмещённости оценки $\bar{x}$ заключаем, что нулевую гипотезу можно записать и так:

$$H_0: \quad M(\bar{\xi}) = a_0.$$

Другими словами, требуется проверить, значимо или нет отличается выборочное среднее от заданного значения. В качестве критерия

Беркович А. МГЭА, кафедра высшей математики 2

выберем величину

$$U = \frac{\bar{x} - a_0}{\sigma/\sqrt{n}} = \frac{\bar{x} - a_0}{\sigma} \cdot \sqrt{n}. \tag{7.28}$$

Аналогичному тому, как это сделано в п. 98.3, можно доказать (сделайте это самостоятельно), что соответствующая случайная величина $U = \dfrac{(\bar{\xi} - a_0)}{\sqrt{n}/\sigma}$ имеет стандартное нормальное распределение. Поэтому в зависимости от конкурирующей гипотезы, решающее правило будет следующим:

- $H_0: \quad M(\xi) = a_0; \qquad H_1: \quad M(\xi) \neq a_0.$

  Из уравнения (7.25) по таблице приложения 2 (или с помощью ЭВМ) определяем $Z_{\text{кр}}$, по формуле (7.28) находим $U_{\text{набл}}$ для имеющихся наблюдений.

  Если $|U_{\text{набл}}| > Z_{\text{кр}}$, гипотезу H_0 отвергаем с уровнем значимости α, если $|U_{\text{набл}}| \leqslant Z_{\text{кр}}$, то нет оснований отвергнуть гипотезу H_0 в пользу данной гипотезы H_1.
- $H_0: \quad M(\xi) = a_0; \qquad H_2: \quad M(\xi) > a_0.$

  Из уравнения (7.26) определяем $Z_{\text{кр}}$, по формуле (7.28) находим $U_{\text{набл}}$. Если $U_{\text{набл}} > Z_{\text{кр}}$, гипотезу H_0 отвергаем с уровнем значимости α, если $U_{\text{набл}} \leqslant Z_{\text{кр}}$, то нет оснований отвергнуть H_0.
- $H_0: \quad M(\xi) = a_0; \qquad H_3: \quad M(\xi) < a_0.$

  Из уравнения (7.26) определяем $Z_{\text{кр}}$, по формуле (7.28) находим $U_{\text{набл}}$. Если $U_{\text{набл}} < -Z_{\text{кр}}$, гипотезу H_0 отвергаем с уровнем значимости α, если $U_{\text{набл}} \geqslant -Z_{\text{кр}}$, то нет оснований отвергнуть H_0.

Если в условиях п. 98.4 дисперсия неизвестна, в качестве критерия следует выбрать величину

$$T = \frac{\bar{x} - a_0}{S^*/\sqrt{n}} = \frac{\bar{x} - a_0}{S^*} \cdot \sqrt{n}. \tag{7.29}$$

Можно доказать (мы не будем этого делать), что соответствующая случайная величина $T = (\bar{\xi} - a_0) \cdot \sqrt{n}/S^*$ имеет распределение Стьюдента с $n-1$ степенью свободы. Решающее правило в зависимости от конкурирующей гипотезы будет следующим:

- $H_0: \quad M(\xi) = a_0; \qquad H_1: \quad M(\xi) \neq a_0.$

  Критическая область в данном случае будет двусторонней; критическая точка t_2 определяется по заданным α и $n-1$ по

Беркович А. МИРЭА, кафедра высшая математика 2

специальным таблицам (приложение 3) или с помощью функции, обратной к функции распределения Стьюдента, имеющейся, например, среди статистических функций Excel. По формуле (7.29) определяем $T_{\text{набл}}$.

Если $|T_{\text{набл}}| > t_2$, гипотеза H_0 отвергается с уровнем значимости α, если $|T_{\text{набл}}| \leqslant t_2$, то нет оснований отвергнуть H_0 в пользу данной гипотезы H_1.

При конкурирующих гипотезах $H_2: \ M(\xi) > a_0$ и $H_3: \ M(\xi) < a_0$ строят соответственно правостороннюю и левостороннюю критические области (см. [5]).

7.13. Сравнение вероятности с заданными значением

Пусть проведено n независимых испытаний Бернулли с неизвестной вероятностью p появления события A в каждом. По результатам испытаний найдена относительная частота m/n, где m — число появлений события A в n испытаниях. Требуется по величине m/n с уровнем значимости α проверить нулевую гипотезу H_0 о том, что неизвестная вероятность p равна заданному значению p_0:

$$H_0: \quad p = p_0.$$

Заметим, что в силу несмещённости оценки m/n для p нулевую гипотезу можно записать и так:

$$H_0: M\Big(\frac{m}{n}\Big) = p_0.$$

Другими словами, требуется проверить, значимо или нет отличается частота от значений p_0. В качестве критерия проверки гипотезы примем величину

$$U = \frac{\frac{m}{n} - p_0}{\sqrt{p_0 q_0}} \cdot \sqrt{n}, \qquad \text{где } q_0 = 1 - p_0. \tag{7.30}$$

Соответствующая случайная величина при справедливости гипотезы H_0 имеет стандартное нормальное распределение. При этом рассуждения аналогичны приведённым в п. 98.3 для случая известной дисперсии, с учётом того, что $M\Big(\frac{m}{n}\Big) = p_0$, $D\Big(\frac{m}{n}\Big) = \frac{p_0 q_0}{n}$.

В зависимости от конкурирующей гипотезы решающее правило будет таким же, как в п. 98.4 для случая известной дисперсии, но значение $U_{\text{набл}}$, конечно, следует вычислять по формуле (7.30).

Берков Н.А. МИРЭА, кафедра высшая математика 2

7.14. Критерии согласия

Геометрический метод определения вида распределения. Критерий Пирсона. Критерий χ^2

Определение 7.16. *Критериями согласия называют критерии для проверки гипотез о виде закона распределения случайной величины.*

7.15. Геометрический метод определения вида распределения

Рассмотрим сначала наглядный, но не строгий метод определения вида распределения. Предположим, что теоретическая функция распределения $F(x)$ зависит от двух неизвестных параметров α и β; чтобы подчеркнуть это, обозначим её $F(x,\alpha,\beta)$: $F(x)=F(x,\alpha,\beta)$. Общий вид функции $y=F(x,\alpha,\beta)$ считается известным, необходимо оценить значения неизвестных параметров α и β. Основная идея графического метода состоит в выборе такой замены координат $u=u(x), v=v(x)$, чтобы график функции распределения $y=F(x,\alpha,\beta)$ в новых координатах $(u;v)$ стал прямой линией $V=k\cdot u+b$ с коэффициентами k и b, зависящими от неизвестных параметров α и β: $k=f(\alpha,\beta)$, $b=\psi(\alpha,\beta)$. Другими словами, уравнение функции распределения $y=F(x,\alpha,\beta)$ в координатах $(u;v)$ должно принять вид:

$$v=k\cdot v+b \quad \text{или} \quad v=f(\alpha,\beta)\cdot u+\psi(\alpha,\beta). \tag{7.31}$$

Поскольку эмпирическая функция распределения $y=F_{\text{э}}(x)$ при достаточно большом объёме статистики лежит вблизи от теоретической функции распределения $y=F(x,\alpha,\beta)$, то после замены переменных график эмпирической функции распределения в координатах $(u;v)$ должен лежать вблизи прямой (7.31). Новая система координат $(u;v)$ с нанесёнными соответствующими значениями $(x;y)$ называются *вероятностной бумагой* . Построив в координатах $(u;v)$ график эмпирической функции распределения, проводят прямую так, чтобы по обе стороны от неё находилось примерно одинаковое количество «ступенек» графика функции $y=F_{\text{э}}(x)$; затем определяют k-величину тангенса угла, образованного этой прямой с осью Ou, и b-координату пересечения с осью Ov. Приравняв полученные величины к их теоретическим значениям $f(\alpha,\beta)$ и $\psi(\alpha,\beta)$, находят оценки неизвестных

Беляева Н.А., МИРЭА, кафедра высшая математика 2

параметров α и β из системы уравнений:

$$\begin{cases} k = f(\alpha, \beta), \\ b = \psi(\alpha, \beta). \end{cases}$$

Если неизвестный параметр один, для нахождения оценки достаточно одного уравнения.

Для нормального закона распределения

$$F(x, m, \sigma) = \Phi\Big(\frac{x-m}{\sigma}\Big) + 0{,}5,$$

где $\Phi(x) = \dfrac{1}{\sqrt{2\pi}} \displaystyle\int_0^x e^{-\frac{z^2}{2}} dz$ — функция Лапласа, используем следующую замену координат:

$$u = x, \quad v = \Phi^{-1}(y - 0{,}5), \tag{7.32}$$

где $x = \Phi^{-1}(y)$ — функция, обратная к функции Лапласа. В новых координатах $(u; v)$ уравнение функции распределения $y = \Phi(\frac{x-m}{\sigma}) + 0{,}5$ примет следующий вид:

$$v = \Phi^{-1}(y - 0{,}5) = \Phi^{-1}\left(\Phi\Big(\frac{x-m}{\sigma}\Big) + 0{,}5 - 0{,}5\right) =$$

$$= \Phi^{-1}\left(\Phi\Big(\frac{x-m}{\sigma}\Big)\right) = \frac{x-m}{\sigma} = \frac{u-m}{\sigma}.$$

Получившееся уравнение:

$$v = \frac{u-m}{\sigma} \tag{7.33}$$

есть уравнение прямой линии вида (7.31), где

$$f(m; \sigma) = \frac{1}{\sigma}, \quad \psi(m; \sigma) = -\frac{m}{\sigma}.$$

Оценив по графику эмпирической функции распределения (7.33) величины k (угловой коэффициент) и b (пересечение с осью Ov), значения m и b получим из системы:

$$\begin{cases} k = \dfrac{1}{\sigma}, \\ b = -\dfrac{m}{\sigma}, \end{cases} \iff \begin{cases} \sigma = \dfrac{1}{k}, \\ m = -\dfrac{b}{k}. \end{cases} \tag{7.34}$$

Берков Н.А. МИРЭА, кафедра высшая математика 2

Для удобства построения графика эмпирической функции распределения (7.33) можно воспользоваться обычной миллиметровой бумагой или специальной «нормальной вероятностной бумагой» с нелинейным масштабом по вертикальной оси, где около значений $v = \Phi^{-1}(y - 0{,}5)$ отмечаются соответствующие значения y.

Пример 7.10. *На рис. 30 для приведённых в табл. 7.1 — 7.3 данных в координатах (u, v) (формулы (7.32)) изображены графики эмпирических функций распределения (ступенчатые функции I — III) и аппроксимирующие их графики теоретических распределений (непрерывные линии). Для наглядности значения $F_э(x)$ в табл. 7.2 приведены в виде простых дробей, в табл. 7.1, 7.3 эти же значения приведены в виде десятичных дробей. Параметры m и σ оцениваются с помощью соотношения (7.34) исходя из величин k и b, которые определяются по графикам в координатах (u, v) с учётом масштаба по осям Ou и Ov.*

Для распределения (I): $k = 1,0$; $\quad b = -1,8 \Longleftrightarrow m = 1,8$; $\sigma = 1,0$;

для распределения (II): $k = 1,4$; $\quad b = -5,7 \Longleftrightarrow m = 4,1$; $\sigma = 0{,}71$;

график распределения (III) имеет заметную искривлённость, что говорит о том, что теоретическая функция распределения (III) не является нормальной.

Для экспоненциального распределения:

$$y = F(x, \lambda) = \begin{cases} 1 - e^{-\lambda x} & x \geqslant 0, \\ 0 & x < 0 \end{cases}$$

используют следующую замену координат:

$$u = x, \qquad v = -\ln(1 - y) \tag{7.35}$$

В новых координатах уравнение теоретической функции распределения $y = F(x, \lambda)$ примет следующий вид:

$$v = -\ln(1 - y) = -\ln\Big(1 - F(x, \lambda)\Big) = -\ln\Big(1 - (1 - e^{-\lambda x})\Big) = \\ = -\ln(e^{-\lambda x}) = \lambda x = \lambda u.$$

Получившееся уравнение $v = \lambda \cdot u$ и есть уравнение прямой линии вида (7.31), где $k = \lambda$, $\quad b = 0$; неизвестный параметр λ равен тангенсу угла наклона аппроксимирующей прямой.

Беляев Н.А. МИРЭА, кафедра высшей математики 2

Вариационные ряды наблюдений

Таблица 7.1 (I)

i	x_i	$y = F_э(x)$
1	0,48	0,07
2	0,55	0,13
3	0,76	0,20
4	0,83	0,27
5	1,34	0,33
6	1,39	0,40
7	1,39	0,47
8	1,94	0,53
9	2,05	0,60
10	2,24	0,67
11	2,52	0,73
12	2,71	0,80
13	2,81	0,87
14	3,44	0,93
15	4,52	1,00

Таблица 7.2 (II)

x_i	$y = F_э(x)$
2,98	1/15
3,03	2/15
3,17	3/15
3,22	4/15
3,57	5/15
3,59	6/15
3,95	7/15
3,96	8/15
4,03	9/15
4,16	10/15
4,35	11/15
4,47	12/15
4,54	13/15
4,96	14/15
5,01	1

Таблица 7.3 (III)

x_i	$y = F_э(x)$
1,68	0,07
1,83	0,13
2,29	0,20
2,46	0,27
4,02	0,33
4,19	0,40
6,54	0,47
6,64	0,53
7,17	0,60
8,30	0,67
10,08	0,73
11,46	0,80
12,21	0,87
17,78	0,93
18,60	1,00

Пример 7.11. *На рис. 31 для приведённых в табл. 7.3 данных изображены графики эмпирической функции распределения и аппроксимирующей прямой в координатах* $(u; v)$ *(формулы (7.35)). Из графика видно:* $k = 0{,}12 \Longleftrightarrow \lambda = 0{,}12$. *Для удобства построения графиков на вертикальной оси* Ov *нанесены соответствующие* v *значения* y *по формуле:* $v = -\ln(1 - y)$.

7.16. Критерий Пирсона проверки гипотезы о виде закона распределения

Пусть имеется случайная выборка, состоящая из n элементов. Требуется найти закон распределения изучаемой случайной величины ξ (или, как условились говорить, генеральной совокупности), определить его параметры и оценить согласие выборки с принятым законом распределения.

На основании статистического материала проверяется гипотеза H_0, состоящая в том, что случайная величина ξ подчиняется некоторому

Бегляв Н.А. МИРЭА, кафедра высшая математика 2

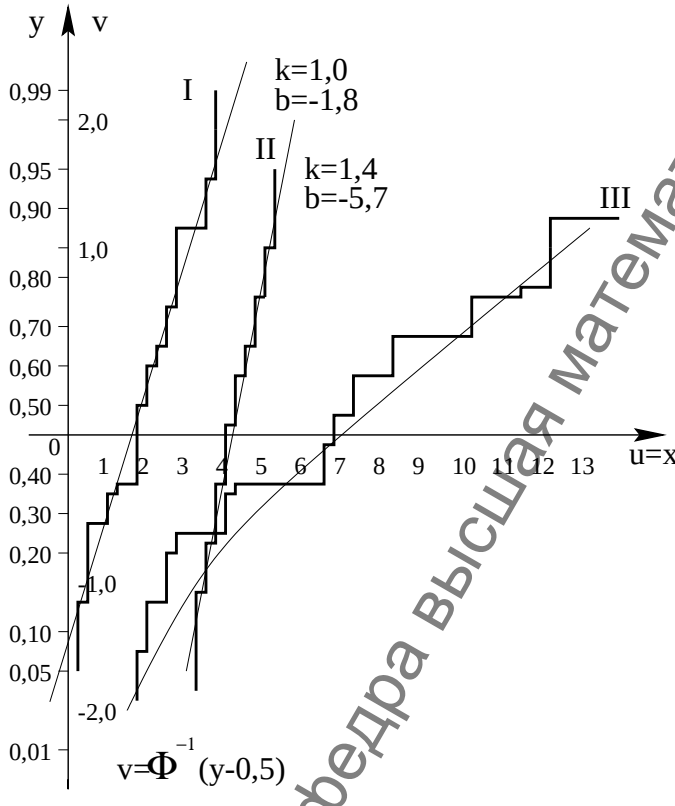

Рис. 30. *Вероятностная бумага для нормального распределения*

закону распределения. Для того чтобы принять или отвергнуть гипотезу H_0, рассматривается величина U — степень расхождения теоретического и статистического распределения. За U принимают сумму квадратов (с некоторыми коэффициентами) отклонений теоретических вероятностей P_i от соответствующих частот P_i^* (критерий χ^2).

Схема расчётов с помощью критерия Пирсона (критерия χ^2) следующая.

(1) На основании выборки выбираем в качестве предполагаемого какой–то закон распределения изучаемой величины (например, с помощью вероятностной бумаги) и оцениваем его параметры, как описано выше.

(2) Всё множество наблюдений разбиваем на s интервалов вида $(a_{j-1}; a_j]$ и подсчитываем эмпирические частоты — количество

Беляев Н.А. МИРЭА, кафедра высшая математика 2

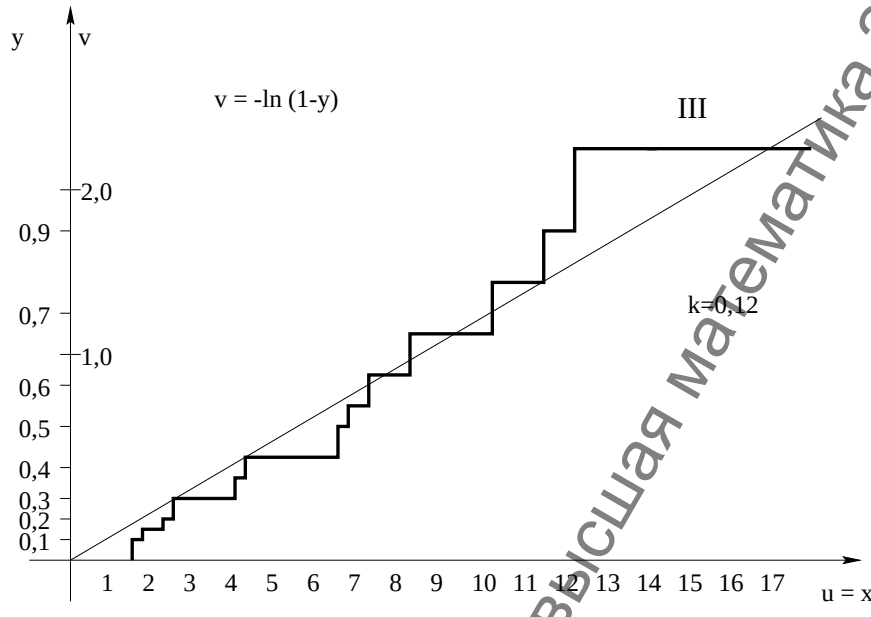

Рис. 31. *Вероятностная бумага для экспоненциального распределения*

наблюдений m_j, попавших в j-ый интервал (см. п. 96.3). Относительная частота наблюдений, попавших в j-ый интервал, равна $P_j^* = \dfrac{m_j}{n}$, $(m_1 + \ldots + m_s = n)$, сумма всех частот, очевидно, равна единице.

(3) Определяем теоретические частоты m'_j для j-го интервала $(a_{j-1}; a_j]$:

$$m'_j = \Big(F(a_j) - F(a_{j-1})\Big) \cdot n,$$

где $F(x)$ – теоретическая функция распределения, найденная на этапе 1.

(4) Вычисляем критерий $\chi^2_{\text{набл}}$ (критерий Пирсона):

$$\chi^2_{\text{набл}} = \sum_{j=1}^{S} \frac{(m_j - m'_j)^2}{m'_j}. \tag{7.36}$$

Из этого выражения видно, что $\chi^2_{\text{набл}}$ равно нулю лишь при совпадении всех соответствующих эмпирических и теоретических частот: $m_i = m'_i$ $(i = 1, 2, \ldots, l)$. В противном

Берков Н.А. МИРЭА, кафедра высшая математика 2

случае $\chi^2_{\text{набл}}$ отлично от нуля и тем больше, чем больше расхождение между частотами. Величина χ^2, определяемая равенством (7.36), является случайной, и (при больших n) имеет χ^2 — распределение с k степенями свободы (принимается без доказательства).

(5) Определяем число степеней свободы k случайной величины χ^2:

$$k = s - 1 - r, \tag{7.37}$$

где r — число параметров закона распределения (для нормального закона распределения $r = 2$), s — число интервалов.

(6) По заданному уровню значимости α и числу степеней свободы k по таблице критических точек распределения χ^2 (таблица приложения 4) находим критическую точку $\chi^2_{\text{кр}}(\alpha; k)$. Если $\chi^2_{\text{набл}} < \chi^2_{\text{кр}}(\alpha; k)$ — нет оснований отвергнуть гипотезу о принятом (нормальном) законе распределения. Если $\chi^2_{\text{набл}} > \chi^2_{\text{кр}}(\alpha; k)$ — гипотезу отвергают с уровнем значимости α.

Пример 7.12. *С помощью критерия Пирсона проверить гипотезу о нормальном распределении выборки, представленной в таблице 7.2 примера 7.10.*

Р е ш е н и е: Разобьём всё множество значений выборки табл. 7.2 на 6 интервалов, границы которых занесены во второй столбец табл. 7.4.

Таблица 7.4

Решение примера 7.2				
j	a_j	m_j	$F(a_j)$	m'_j
0	2,5	1	0,0155	0,969
1	3,0	3	0,0800	2,659
2	3,5	4	0,2573	4,246
3	4,0	4	0,5404	3,948
4	4,5	2	0,8036	2,137
5	5,0	1	0,9460	0,673
6	5,5		0,9909	

В третий столбец табл. 7.4 заносим количество наблюдений m_j, попавших в j-ый интервал. По формулам (7.2), (7.12), (7.5) определяем параметры нормального распределения $\bar{x}$ и S^* для выборки из табл. 7.2:

$$\bar{x} = 3,933; \quad S^* = 0,664$$

Беркинов Н.А. МИРЭА, кафедра высшая математика 2

и находим значения теоретической функции распределения $F(a_j)$. В данном примере $F(a_j) = \Phi\Big(\dfrac{a_j - \bar{x}}{S^*}\Big) + 0{,}5$. В пятый столбец заносим теоретические частоты m'_j, вычисляемые, как указано выше.

По формуле (7.36) находим значение $\chi^2_{\text{набл}} = 0{,}228$. По таблице приложения 4 для $\alpha = 0{,}05$ и $k = 6 - 1 - 2 = 3$ находим критическую точку $\chi^2_{\text{кр}}(0{,}05; 3) = 7{,}8$. Поскольку $\chi^2_{\text{набл}} < \chi^2_{\text{кр}}(0{,}05; 3)$, нет оснований отвергать гипотезу H_0 о нормальном распределении выборки из таблицы 7.2. Заметим, что этот результат хорошо согласуется с данными вероятностной бумаги (график I на рис. 30).

Покажем использование критерия Пирсона с применением программ в рамках пакета Maxima.

Пример 7.13. *Для исследования вида некоторой зависимости произведено 100 испытаний. Результаты полученных испытаний разбили на 9 диапазонов, границы которых занесены в массив a:*

На практике чаще всего выборка большого объёма записывается в текстовый файл. Затем данные считываются программой обработки, и полученная выборка анализируется. Поэтому мы также разобьем задачу на две подзадачи. В первой сгенерируем выборку, а во второй её обработаем. Для генерации выборки используем команду random_normal(M, σ, n), возвращающую список из n псевдослучайных чисел, близких к нормальному закону распределения с математическим ожиданием M и среднеквадратическим отклонением σ. Полученный список $x1$ запишем в файл pirson.txt, находящийся на диске D в папке mymaxima.

/* Первая программа, генерирующая выборку объёма n=100 и записывающая её в текстовый файл "D:/mymaxima/pirson.txt").*/

```
(%i1) n:100$   fpprintprec:3$
(%i3) load(distrib)$
(%i4) x1:random_normal(120, 25, n)$
(%i5) write_data(x1, "D:/mymaxima/pirson.txt")$
```

Пример 7.14. *Для исследования вида некоторой зависимости произведено 100 испытаний, результаты которых записаны в текстовый файл "D:/mymaxima/pirson.txt"). С помощью критерия Пирсона проверить гипотезу о нормальном распределении полученной выборки.*

Maxima-программа:

Беркович А. МИРЭА, кафедра высшая математика 2

```
(%i1) fpprintprec:3$ n:100$ numer:true$
(%i4) load(distrib)$
```

/ Считываем выборку в список x1.*/*

```
(%i5) x1:read_list("D:/mymaxima/pirson.txt");
```

*/*Сортируем список x1 в порядке возрастания значений.*/*

```
(%i6)  x2:sort(x1);
```

/ Разбиваем выборку на s интервалов постоянной длины delta.*/*

```
(%i7) s:9; delta:(x2[n] -x2[1])/s;
```

*/*Координаты границ элементов.*/*

```
(%i9) a:makelist(x2[1]+(i-1)*delta, i, 1, s+1);
(%o9) [66.9, 79.6, 92.3, 105., 118., 130., 143., 156.,
       168., 181.]
```

/ Координаты середин элементов.*/*

```
(%i10)  U:makelist((a[j]+a[j+1])/2, j, 1, s);
(%o10) [73.2, 85.9,  98.6, 111., 124., 137., 149., 162., 175.]
```

*/*Определение частоты наблюдений по интервалам.*/*

```
(%i11) m:makelist(0, i, 1, s); for j:1 while j<=n do(
    k:fix((x2[j] -x2[1])/delta)+1,if k>s then k:s,m[k]:m[k]+1);
```

/ Вывод значений эмпирической частоты наблюдений по интервалам.*/*

```
(i13) m;
(%o13) [3, 9, 23, 11, 19, 16, 11, 5, 3]
```

/ Контроль объёма выборки.*/*

```
(%i14) sum(m[i], i, 1, s);
(%o14) 100
```

/ Строим график.*/*

Берлов Н.А. МИРЭА, кафедра высшая математика 2

```
(%i15) wxplot2d([['discrete,makelist([U[j], m[j]], j, 1, s)]],
      [style,[lines, 3, 5]],  [gnuplot_preamble,"set  grid"],
      [ylabel,""])$
(%i16) Mx:sum(m[j]*U[j], j, 1, s)/n;
(%o16) 117.
(%i17) S2:n/(n-1)*(sum(m[j]*U[j]^2, j, 1, s)/n-Mx^2);
(%o17) 648.
(%i18) S:sqrt(S2);
(%o18) 25.5
(%i19) F(x):=cdf_normal (x, 0, 1) -0.5;
(%o19) F(x):=cdf_normal(x, 0, 1) -0.5
```

/**Теоретические вероятности.**/

```
(%i20) P:makelist(F((a[j+1] -Mx)/S) -F((a[j] -Mx)/S), j, 1, s);
         P[1]:F((a[2] -Mx)/S)+0.5;
        P[s]:0.5-F((a[s] -Mx)/S);
(%o20) [0.006, 0.03, 0.09, 0.2, 0.2, 0.2, 0.1, 0.06, 0.02]
(%o21) 0.007
(%o22) 0.02
```

/**Теоретические частоты.**/

```
(%i23) m1:makelist(n*P[j], j, 1, s);
(%o23) [0.7, 2.81, 8.55, 17.6, 24.4, 22.9, 14.6, 6.25, 2.22]
```

/**Контроль объёма выборки для теоретических частот.**/

```
(%i24) sum(m1[j], j, 1, s);
(%o24) 100
```

/* *Вычисление* $\chi^2_{\text{набл}}$. */

```
(%i25) x2nabl:sum((m[j] -m1[j])^2/m1[j], j, 1, s);
(%o25) 5.69
```

Так как $\chi^2_{\text{набл}}$ =5,69 наблюдаемое меньше критического равного, $\chi^2_{\text{кр}}$=12,6, то, нет основания отвергать выдвинутую гипотезу о нормальном распределении исследуемой выборки объёма $n = 100$.

Ответ: На базе полученной выборки делаем вывод, что исследуемая непрерывная случайная величина подчиняется нормальному закону распределения.

Беркович Л.А. МИРЭА, кафедра Высшая математика 2