ЛАБОРАТОРНАЯ РАБОТА №3

ОЦЕНКА ЧИСЛОВЫХ ХАРАКТЕРИСТИК СЛУЧАЙНЫХ ВЕЛИЧИН

1. ЦЕЛЬ РАБОТЫ

1. Изучить методы нахождения числовых характеристик случайных величин (с.в.)

2. Произвести экспериментальные исследования зависимости точности оценок числовых характеристик от объема выборки случайной величины.

2. ТЕОРЕТИЧЕСКИЙ РАЗДЕЛ

Неслучайные параметры, выражающие в сжатой форме наиболее существенные особенности распределения *случайной величины*, называются ее *числовыми характеристиками.* Эти числовые характеристики находятся, как правило, путем осреднения по всему числу испытаний некоторых неслучайных функций исследуемой с.в.

Допустим, что с.в. ξ в *j*-м испытании приняла конкретное значение, x_j^* и полное число этих испытаний есть *N*. Тогда среднее арифметическое величины ξ, обозначаемое как $\widetilde{M}_1$, есть

$$\widetilde{M}_1 = \frac{1}{N}\sum_{j=1}^{N} x_j^* . \qquad (2.1)$$

Эта величина случайна, однако при $N \to \infty$ она в силу статистической устойчивости стремится к некоторому пределу, носящему название *математического ожидания* (МО) величины ξ. Оно обозначается как M_1. Для дискретной с.в. оно выражается формулой

$$M_1 = M[\xi] = \sum_{i=1}^{n} x_i P_i , \qquad (2.2)$$

а для непрерывной – формулой

$$M_1 = M[\xi] = \int_{-\infty}^{\infty} x\, p(x) dx . \qquad (2.3)$$

В формуле (2.2) *n* величин x_i представляют собой полную совокупность значений, которые может принимать дискретная с.в. ξ, а P_i - вероятности этих значений. В формуле (2.3) $p(x)$ есть плотность вероятности непрерывной с.в. ξ.

Строго говоря, $\widetilde{M}_1$ не совпадает с M_1, и это совпадение достигается только при $N \to \infty$. Следовательно, точное значение МО может быть найдено по формулам (2.2) и (2.3) при точном знании P_i или $p(x)$, которые не всегда известны. В то же время экспериментально-расчетным путем по формуле (2.1) может быть найдено только его приближенное значение $\widetilde{M}_1$, которое в связи с этим называется *оценкой* математического ожидания.

Итак, в силу данных выше определений M_1 является числовой характеристикой с.в., а $\widetilde{M}_1$ - ее приближенной оценкой. Величина M_1 определяет некоторую среднюю величину ξ, вокруг которого группируются ее все возможные значения.

Другие числовые характеристики с.в. ξ находятся путем осреднения некоторых детерминированных функций случайного аргумента $\varphi(\xi)$. Если число испытаний, конечно, то по аналогии с формулой (2.1) получим оценки таких характеристик в виде

$$\widetilde{\varphi}(\xi)=\frac{1}{N}\sum_{j=1}^{N}\varphi\left(x_j^*\right). \tag{2.4}$$

При $N\to\infty$ они переходят в МО этих функций:

$$M[\varphi(\xi)]=\sum_{i=1}^{n}\varphi(x_i)P_i \tag{2.5}$$

для дискретной с.в. ξ и

$$M[\varphi(\xi)]=\int_{-\infty}^{\infty}\varphi(x)p(x)dx \tag{2.6}$$

для непрерывной.

На практике наибольшую применимость имеют *центральные моменты* различных порядков, обозначаемые как μ_k. Для них $\varphi(\xi)=(\xi-M_1)^k$, где порядок *k* – целые неотрицательные числа. Величина $(\xi-M_1)$, получаемая из каждого значения *исходной* с.в. ξ вычитанием ее МО, называется *центрированной*, а сама процедура этого вычитания – *центрированием*. Итак, имеем оценку момента *k*-го порядка

$$\widetilde{\mu}_k=\frac{1}{N}\sum_{j=1}^{N}\left(x_j^*-M_1\right)^k. \tag{2.7}$$

При $N\to\infty$ отсюда получим

$$\mu_k=\sum_{i=1}^{n}(x_i-M_1)^kP_i \tag{2.8}$$

для дискретной с.в. и

$$\mu_k=\int_{-\infty}^{\infty}(x-M_1)^k p(x)\,dx \tag{2.9}$$

для непрерывной с.в.

Физическая размерность μ_k и $\widetilde{\mu}_k$ есть

$$[\mu_k] = [\widetilde{\mu}_k] = [\xi]^k \quad (2.10)$$

Центральный момент *второго порядка*

$$\sigma^2 = \mu_2 \quad (2.11)$$

называется *дисперсией* с.в. ξ, а квадратный корень из нее σ - *среднеквадратическим отклонением* с.в. ξ. Величина

$$\widetilde{\sigma}^2 = \widetilde{\mu}_2 \quad (2.12)$$

есть *оценка* этой дисперсии, а $\widetilde{\sigma} = \sqrt{\widetilde{\sigma}^2}$ - оценка среднеквадратического значения с.в. ξ. Величина σ характеризует полуширину распределения вероятности или плотности распределения вероятности.

Нетрудно показать, что центральный момент третьего порядка μ_3 равен нулю, если распределение симметрично относительно своего МО, и отличен от нуля в противном случае. Однако применять его непосредственно для оценки степени асимметрии распределения неудобно, так как он имеет размерность $[\xi]^3$. Для этого применяют *безразмерную* величину

$$\gamma_1 = \mu_3 / \sigma^3, \quad (2.13)$$

называемую *коэффициентом асимметрии* величины ξ. Этот коэффициент характеризует *скошенность* распределения или плотности распределения вероятности. Одновершинное распределение с $\gamma_1 < 0$ имеет левостороннюю (отрицательную) асимметрию, т.е. распределение имеет слева «хвост». Если $\gamma_1 > 0$, оно имеет «хвост» справа. Для симметричного распределения $\gamma_1 = 0$.

Безразмерная величина

$$\gamma_2 = \frac{\mu_4}{\sigma^4} - 3 \quad (2.14)$$

называется *коэффициентом эксцесса* распределения и характеризует степень его *островершинности* в сравнении с нормальным (гауссовским) распределением. Для нормального распределения эта величина равна нулю. Для более островершинного распределения $\gamma_2 > 0$. Для менее островершинного $\gamma_2 < 0$. При этом сравнении необходимо считать, что у всех рассматриваемых распределений величина σ^2 одинакова.

Статистический пакет Statistics & Machine Learning Toolbox системы MATLAB поддерживает 20 видов распределений вероятности: 14 непрерывных и 6 дискретных (таблица 2.1).

В этой таблице: A, B, MU, NU, NU1, NU2, V1, V2, DELTA, LAMBDA, NN, M, K, P, RR – параметры, описывающие распределения; R – матрица размером m×n, состоящая из случайных величин ξ, имеющих указанное распределение; M –математическое ожидание $M[\xi]$ и V – дисперсия с.в. ξ. Команда из столбца IV дает возможность вычислить теоретическое значение МО $M_1 = M[\xi]$, обозначаемое здесь как M, и теоретическую дисперсию σ^2, обозначаемую как V.

Входящий в MATLAB пакет Statistics Toolbox имеет в своем составе демонстрационные программы, создающие интерактивную среду для генерации случайных чисел, изучения их различных распределений вероятностей и других целей.

Таблица 2.1 – Распределение вероятностей, команды генерации случайных величин и команды вычисления их числовых характеристик

Вариан	Вид распределения	Команда генерации случайной величины	Команда вычисления M_1 и σ^2
Непрерывные распределения			
1	Бета	R=betarnd(A,B,m,n)	[M,V]=betastat(a,b)
2	Экспоненциальное	R=exprnd(MU,m,n)	[M,V]=expstat(MU)
3	Гамма	R=gamrnd(A,B,m,n)	[M,N]=gamstat(A,B)
4	Логнормальное	R=lognrnd(MU,SIGMA,m,n)	[M,V]=lognstat(MU,SIGMA)
5	Нормальное (гауссовское)	R=normrnd(MU,SIGMA,m,n)	[M,V]=normstat(MU,SIGMA)
6	Релея	R=raylrnd(B,m,n)	[M,V]=raylstat(B)
7	Равномерное	R=unifrnd(A,B,m,n)	[M,V]=unifstat(A,B)
8	Вейбулла	R=weibrnd(A,B,m,n)	[M,V]=weibstat(A,B)
9	Хи-квадрат	R=chi2rnd(NU,m,n)	[M,V]=chi2stat(NU)
10	Нецентральное хи-квадрат	R=ncx2rnd(NU,DELTA,m,n)	[M,V]=ncx2stat(NU,DELTA)
11	Фишера-Снекора (F-распредел.)	R=frnd(V1,V2,m,n)	[M,V]=fstat(V1,V2)
12	Нецентральное F-распределение	R=ncfrnd(NU1,NU2,DELTA,m,n)	[M,V]=ncfstat(NU1,NU2,DELTA)
13	Стьюдента (t-распределение)	R=trnd(NU,m,n)	[M,V]=tstat(NU)
14	Нецентральное t-распределение	R=nctrnd(NU,DELTA,m,n)	[M,V]=nct(NU,DELTA)
Дискретные распределения			
	Биномиальное	R=binornd(NN,P,m,n)	[M,V]=binostat(NN,P)
	Равновероятное	R=unidrnd(NN,m,n)	[M,V]=unidstat(NN)
	Геометрическое	R=geornd(P,m,n)	[M,V]=geostat(P)
	Гипергеометрическое	R=hygernd(M,K,NN,m,n)	[M,V]=hygestat(M,K,NN)
	Отрицательное	R=nbinrnd(RR,P,m,n)	[M,V]=nbinrnd(RR,P)

	биномиальное		
	Пуассона	R=poissrnd(LAMBDA,m,n)	[M,V]=poisstat(LAMBDA)

Так, оператор **disttool**, введенный в командном окне MATLAB, открывает окно, в котором изображены кривые теоретических зависимостей любого из имеющихся в MATLAB распределений. Последние могут быть выбраны в выпадающем меню в середине верхней части этого окна. Если вверху в правой части окна выбрать *pdf* (probability density function), на графике будет изображена кривая *плотности* вероятности $p(x)$ рассматриваемой *непрерывной* с.в. или набор *значений вероятности* P_i *дискретной* с.в. Если же выбрать *cdf* (cumulative distribution function), то отобразится *интегральная функция распределения* $F(x)$ данной с.в.

Если в командном окне ввести оператор **randtool**, то откроется окно, в котором в виде гистограммы будет продемонстрировано эмпирическое распределение данной случайной величины при заданном (вверху справа) числе ее отсчетов.

3. ХОД РАБОТЫ

1. Получить у преподавателя вариант задания (Таблица 3.1). Во всех заданиях положить $m=1$ и считать *n* текущим, изменяющимся от 1 до 1000.

2. Написать в системе MATLAB коды для вычисления оценок моментов $\widetilde{M}_1$, $\widetilde{\mu}_1$, $\sigma^2 = \widetilde{\mu}_2$, $\widetilde{\mu}_3$, $\widetilde{\mu}_4$, оценки коэффициента асимметрии

$$\widetilde{\gamma}_1 = \frac{\widetilde{\mu}_3}{\sqrt{\widetilde{\mu}_2^3}} \tag{3.1}$$

и оценки коэффициента эксцесса

$$\widetilde{\gamma}_2 = \frac{\widetilde{\mu}_4}{\widetilde{\mu}_2^2} - 3\,. \tag{3.2}$$

3. С помощью этих кодов рассчитать зависимости указанных оценок от числа испытаний *N*
для 1 ≤ *N* ≤ 1000 и изобразить их графически в линейном и полулогарифмическом (по оси *x*) масштабах. Рисунки снабдить обозначениями переменных по осям и подрисуночными подписями.

4. Найти теоретические значения M_1 и σ^2 и сравнить их с экспериментальными.

5. Применив, оператор **disttool**, установить вид теоретических кривых, характеризующих закон распределения данного варианта случайной величины. Распечатать соответствующие графики.

6. Применив оператор **randtool**, проследить, как меняются эмпирические распределения данной с.в. при последовательном выборе ее числа отсчетов

N=100, 200, 500, 1000. Распечатать соответствующие графики.

Вид распределения	Вар.	Параметры распределения
Нормальное	1.	MU=3 , SIGMA=3
Рэлея	2.	B=2
	3.	B=0,5
Равномерное непрерывное	4.	A=0 , B=2
Вейбулла	5.	A=0,5 , B=2
Хи-квадрат	6.	NU=3
Нецентральное хи-квалрат	7.	NU= 4 , DELTA=2
F-распределение	8.	V1= 6 , V2=8
Нецентральное F-распределение	9.	NU1=10 , NU2=100 , DELTA=4
Стьюдента	10.	NU=7
Нецентральное t-распределение	11.	NU=10 , DELTA=1
Биномиальное	12.	NT=10 , P= 0,5
Равновероятное дискретное	13.	NN=100
Геометрическое	14.	P=0,0003
Гипергеометрическое	15.	M= 1000 , K= 60 NN=30
Отрицательное биномиальное	16.	RR= 3 , P=0,01
Пуассона	17.	LAMBDA=6
Бета	18.	A= 2 , B= 4
Экспоненциальное	19.	MU= 3
Гамма	20.	A=5 , B=7
Логнормальное	21.	MU= 0,5 , SIGMA= 0,25

7. Дать *письменное* объяснение всем наблюдаемым зависимостям.
8. Оформить отчет.

Таблица 3.1 – Варианты задания

4. СОДЕРЖАНИЕ ОТЧЁТА

1. Цель работы.
2. Краткое теоретическое введение.
3. Теоретический расчёт математического ожидания и дисперсии для заданного типа распределения.
4. Графики теоретических кривых, характеризующих закон распределения указанного варианта случайной величины (интегральная функция распределения, функция плотности вероятности).
5. Графики эмпирических распределений указанного варианта случайной

величины.

6. Программа на языке MATLAB для практического расчёта числовых характеристик случайной величины.

7. Выводы по работе в развернутом виде, сравнительная характеристика полученных теоретических значений с практическими, дать письменное объяснение всем наблюдаемым зависимостям.

5. КОНТРОЛЬНЫЕ ВОПРОСЫ

1. Что такое числовые характеристики случайных величин?
2. Геометрический смысл математического ожидания.
3. Геометрический смысл дисперсии.
4. Геометрический смысл среднеквадратического отклонения.
5. Геометрический смысл коэффициента асимметрии.
6. Геометрический смысл коэффициента эксцесса.
7. Что такое случайная величина?
8. Каким образом зависят от числа отсчетов случайной величины оценки числовых характеристик?
9. В чем отличие числовых характеристик с.в. от их оценок?
10. Что такое гистограмма?
11. Чем отличаются гистограммы непрерывных и дискретных случайных величин?
12. Какие особенности гистограмм характеризуют найденные числовые характеристики?
13. Что такое начальный момент распределения?
14. Что такое центральный момент распределения?
15. Какой цели служат начальные и центральные моменты распределения?
16. В чем сходство и различие в статистическом описании дискретных и случайных непрерывных величин?