Федеральное агентство по образованию
АМУРСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ
ГОУ ВПО «АмГУ»

УТВЕРЖДАЮ
Зав. кафедрой ОМиИ
______________Г. В. Литовка
«_____»_____________2007

г.

МАТЕМАТИКА

Часть 3

УЧЕБНО-МЕТОДИЧЕСКИЙ КОМПЛЕКС ПО ДИСЦИПЛИНЕ
для специальностей:
080109, 080105, 080102, 080507, 080502, 080504, 080111

Составители: Г. Н. Торопчина,
Г. П. Вохминцева

Благовещенск 2007 г.

Печатается по решению
редакционно-издательского совета
факультета математики и информатики
Амурского государственного
университета

Г. Н. Торопчина, Г.П. Вохминцева
Учебно-методический комплекс по дисциплине «Математика» для студентов очной формы обучения для специальностей: 080109, 080105, 080102, 080507, 080502, 080504, 080111. – Благовещенск: Амурский гос. ун-т, 2007. – с.

Учебно-методический комплекс ориентированы на оказание помощи студентам очной формы обучения по специальностям: 080109, 080105, 080102, 080507, 080502, 080504, 080111 при изучении курса математики.

Пояснительная записка

Курс состоит из двух частей: первая - “Теория вероятностей”, вторая - “Элементы математической статистики”.

Он является частью обязательного курса «Математика» цикла ЕН, изучается студентами второго курса экономических специальностей. Настоящий комплекс рассчитан на 140 часов трудоемкости согласно учебного плана и составлен на основании Государственного образовательного стандарта 2000 года для указанной специальностей.

Предмет курса - изучение вероятностных закономерностей, возникающих при взаимодействии большого числа случайных факторов, массовых однородных случайных явлений в науке и жизни общества, а также математических методов систематизации и использования статистических данных для научных и практических выводов.

Цель курса:

• создание теоретической базы для изучения курсов “Математические методы в экономике”, “Математическое моделирование” и “Эконометрика”;

• развитие и совершенствование навыков логических построений и доказательств;

• разбор конкретных примеров использования математических методов при решении управленческих, экономических и иных задач социально-гуманитарной направленности.

В данном пособии приводятся основные темы указанного раздела программы “Математики”, а также даются рекомендации по их изучению.

Основные задачи курса – овладение теорией в рамках программы курса и выработка навыков использования теоретического материала в решении практических задач.

В результате изучения первой части курса студент должен знать:

- основные понятия и теоремы теории вероятностей;
- основные законы распределения случайных величин.

В результате изучения *первой части* курса студент должен уметь:

- строить вероятностные модели;
- вычислять вероятности случайных событий;
- применять наиболее важные законы распределения случайных величин и их числовые характеристики.

В результате изучения второй части курса студент должен знать**:**

- основные понятия математической статистики;
- методы обработки и анализа статистических данных в зависимости от целей исследования;
- технику проверки гипотез;
- методы корреляционного и регрессионного анализов.

В результате изучения второй части курса студент должен уметь:

- определить генеральную совокупность и исследуемую случайную величину;
- сформулировать математическую постановку задачи;
- собрать экспериментальный материал и сформировать выборку;
- с учетом поставленной задачи, используя методы математической статистики, провести обработку и анализ данных;
- пользоваться расчетными формулами, таблицами, графиками по теории вероятностей.

Итоговая аттестация студентов проводится в виде *экзамена за 3 семестр.* В рамках курса необходимо выполнить две контрольные работы и выполнить и защитить две расчетно-графические работы (для контроля за самостоятельной работой).

Государственный стандарт

Теория вероятностей и математическая статистика. Сущность и условия применимости теории вероятностей. Основные понятия теории вероятностей. Вероятностное пространство. Случайные величины и способы их описания. Модели законов распределения вероятностей, наиболее употребляемые в социально-экономических приложениях. Закон распределения вероятностей для функции от известных случайных величин. Неравенство Чебышева. Закон больших чисел и его следствие. Особая роль нормального распределения: центральная предельная теорема. Цепи Маркова и их использование в моделировании социально-экономических процессов. Статистическое оценивание и проверка гипотез, статистические методы обработки экспериментальных данных.

Содержание курса

Теория вероятностей

Тема 1. Предмет теории вероятностей. Основные понятия.

Предмет теории вероятностей, первоначальные понятия и определения, основные формулы комбинаторики, классическое определение вероятностей.

Тема 2. Сложение и умножение вероятностей.

Теорема сложения вероятностей, условные вероятности, теорема умножения вероятностей, независимые события и их свойства. Вероятность появления хотя бы одного события.

Тема 3. Формулы полной вероятности и Байеса. Схема Бернулли.

Формула полной вероятности, формула Байеса, схема повторных испытаний Бернулли, формула Бернулли. Локальная и интегральная теоремы Муавра – Лапласа, формула Пуассона.

Тема 4. Случайные величины.

Случайная величина. Примеры случайных величин. Виды случайных величин (конечные, дискретные, непрерывные). Закон и таблица распределения конечных и дискретных случайных величин. Функция распределения случайной величины и её свойства. Плотность распределения непрерывной случайной величины и её свойства. Эффект нулевой вероятности. Математическое ожидание как среднее значение случайной величины. Определение математического ожидания для различных видов случайных величин. Определение суммы и произведения случайных величин. Свойства математического ожидания. Дисперсия случайной величины и её свойства. Среднее квадратичное отклонение.

Тема 5. Основные распределения случайных величин.

Биноминальное распределение и его характеристики. Распределение Пуассона и его характеристики. Теорема Пуассона. Нормальное распределение и его характеристики. Локальная и интегральная теоремы Муавра – Лапласа. Показательное распределение и его характеристики. Равномерное распределение и его характеристики.

Тема 6. Закон больших чисел.

Неравенство Чебышева. Теоремы Чебышева и Бернулли.

Тема 7. Система случайных величин.

Случайные векторные величины. Функция и плотность распределения случайной двумерной величины. Корреляционный момент связи двух случайных величин. Коэффициент корреляции.

Тема 8. Элементы теории массового обслуживания. Случайный процесс и его характеристики.

Понятие о случайном процессе со счётным множеством состояний. Поток событий. Простейший поток и его свойства. Нестационарный пуассоновский поток. Марковский случайный процесс. Системы массового обслуживания и их классификация. Установившийся режим обслуживания. Формулы Эрланга.

Тема 9. Показатели эффективности систем массового обслуживания.

Система дифференциальных уравнений для потока и её решение. Системы массового обслуживания с марковскими потоками состояний.

Математическая статистика.

Тема 1. Статистические оценки параметров распределения.

Основные задачи статистики и математической статистики. Выборки. Статистическая обработка результатов наблюдений. Оценки и связанные сними понятия. Точечные оценки вероятности, математического ожидания, дисперсии и их свойства. Метод максимума правдоподобия и его применения для нахождения точечных оценок параметров основных распределений. Понятие доверительных оценок. Построение доверительных интервалов для параметров нормального распределения: случаи, когда один из параметров известен и когда неизвестны оба параметры.

Тема 2. Проверка статистических гипотез.

Постановка задачи проверки гипотез. Критерий оценки и его мощность. Критическая область и область принятия гипотезы. Проверка гипотез о значениях параметров нормального распределения. Проверка гипотез в виде распределения. Критерий Пирсона.

Тема 3. Корреляционный и регрессионный анализ.

Функциональные и корреляционные зависимости случайных величин. Линейная и нелинейная регрессия. Составление уравнений прямых регрессий, метод наименьших квадратов. Статистическая оценка коэффициента корреляции и её свойства. Построение доверительных интервалов для параметров линейной регрессии. Проверка статистической значимости регрессии и адекватности модели регрессии результатам наблюдений.

Тема 4. Основы факторного анализа.

Постановка задачи факторного анализа. Линейная модель. Примеры практического применения факторного анализа.

Темы лекций

1. Предмет теории вероятностей. Случайные события. Алгебра событий. Относительная частота и вероятность случайного события. Полная группа событий. Классическое определение вероятности. Основные свойства вероятности.

2. Теорема сложения вероятностей. Противоположные события. Условные вероятности. Теорема умножения вероятностей. Независимые события. Вероятность появления хотя бы одного события.

3. Формула полной вероятности и формула Байеса.

4. Схема и формула Бернулли. Приближение Пуассона для схемы Бернулли. Теорема Лапласа.

5. Случайные величины. Закон распределения и функция распределения дискретной случайной величины. Биномиальное распределение и распределение Пуассона.

6. Основные числовые характеристики дискретных и непрерывных случайных величин: математическое ожидание, дисперсия и среднее квадратическое отклонение. Их свойства и примеры.

7. Функция распределения и плотность распределения непрерывной случайной величины, их взаимосвязь и свойства. Равномерное распределение вероятностей.

8. Основные числовые характеристики непрерывных случайных величин: математическое ожидание, дисперсия и среднее квадратическое отклонение. Их свойства и примеры.

9. Равномерное распределение вероятностей. Показательное распределение. Функция надежности. Показательный закон надежности.

10. Нормальный закон распределения вероятностей. Нормальная кривая. Правило трех сигм.

11. Система случайных величин. Плотность распределения системы двух случайных величин. Условные законы распределения. Зависимые и независимые случайные величины. Линейная корреляция.

12. Закон больших чисел. Неравенство Чебышева. Теоремы Чебышева и Бернулли.

13. Теория массового обслуживания. Случайные процессы. Поток событий. Цепи Маркова.

14. Основные понятия математической статистики. Генеральная совокупность и выборка. Вариационный ряд, статистический ряд. Группированная выборка. Группированный статистический ряд. Полигон частот. Выборочная функция распределения и гистограмма.

15. Числовые характеристики статистического распределения: выборочное среднее, оценки дисперсии, оценки моды и медианы, оценки начальных и центральных моментов. Статистическое описание и вычисление оценок параметров двумерного случайного вектора.

16. Основные свойства статистических характеристик параметров распределения: несмещенность, состоятельность, эффективность. Несмещенность и состоятельность выборочного среднего как оценки математического ожидания. Смещенность выборочной дисперсии. Пример несмещенной оценки дисперсии. Асимптотически несмещенные оценки. Способы построения оценок: метод наибольшего правдоподобия, метод моментов, метод квантили, метод наименьших квадратов, байесовский подход к получению оценок.

17. Интервальное оценивание неизвестных параметров. Точность оценки, доверительная вероятность (надежность), доверительный интервал. Построение доверительных интервалов для оценки математического ожидания нормального распределения при известной и при неизвестной дисперсии. Доверительные интервалы для оценки среднего квадратического отклонения нормального распределения.

18. Статистическая проверка статистических гипотез. Общие принципы проверки гипотез. Понятия статистической гипотезы (простой и сложной), нулевой и конкурирующей гипотезы, ошибок первого и второго рода, уровня значимости, статистического критерия, критической области, области принятия гипотезы. Наблюдаемое значение критерия. Критические точки. Мощность критерия. Критерии для проверки гипотез о вероятности события, о математическом ожидании, о сравнении двух дисперсий.

Календарный план занятий

Лекции 2ч практические занятия 2ч по специальностям 080507, 080502, 080504, 080111.
Лекции 1ч практические занятия 3ч по специальностям 080109, 080105, 080102.

№ недели	Темы лекций	Содержание практических занятий	Самостоятельная работа
		Контрольная работа по остаточным знаниям (входной контроль)	
1	Основные понятия теории вероятностей	Вероятность случайного события	Выдача РГР №1. Выполнение домашнего задания
2	Аксиомы теории вероятностей	Теоремы сложения и умножения вероятностей (к.р.)	-
3	Основные теоремы (формулы) теории вероятностей	Формулы полной вероятности и Байесса	-
4	Случайные величины. Законы распределения дискретных с. в.	Схема Бернулли	Подготовка к контрольной работе. Повторение теоретического материала
5	Законы	Контрольная работа.	Сдача РГР №1.

	распределения непрерывных случайных величин	Законы распределения дискретных сл.вел.	Выдача РГР №2
6	Числовые характеристики случайных величин.	Законы распределения непрерывных случайных величин.	Выполнение домашнего задания
7	Основные распределения дискретных случайных величин.	Числовые характеристики случайных величин	-
8	Основные распределения непрерывных случайных величин	Распределение Бернулли, Пуассона, геометр. и др. распр. дискретных случ. величин	-
9	Закон больших чисел	Распределение непрерывных случайных величин	-
10	Система случайных величин	Закон больших чисел	-
11	Случайные процессы. Характеристики случайного процесса	Контрольная работа	Сдача РГР №2
12	Цепи Маркова. Марковский случайный процесс. Система массового обслуживания.	Метод наименьших квадратов	Защита лабораторной работы
13	Основные задачи математической статистики. Обработка результатов наблюдений	Однофакторный анализ (л.р.)	Защита лабораторной работы

14	Точечные и интервальные оценки параметров распределения	Точечные оценки параметров распределения (л.р.)	-
15	Проверка статистических гипотез	Интервальные оценки параметров распределения (л.р.)	-
16	Проверка статистических гипотез	Проверка статистических гипотез (л.р.)	-
17	Элементы корреляционного и регрессионного анализа	Проверка статистических гипотез (л.р.)	-
18	Обзор	Элементы теории корреляции (л.р.)	-

КОНСПЕКТ ЛЕКЦИЙ ПО ТЕОРИИ ВЕРОЯТНОСТЕЙ

Лекция 1.

Предмет теории вероятностей. Случайные события. Алгебра событий. Относительная частота и вероятность случайного события. Полная группа событий. Классическое определение вероятности. Основные свойства вероятности.

Основные понятия.

Определение. Событием называется всякий факт, который может произойти или не произойти в результате опыта.

При этом тот или иной результат опыта может быть получен с различной степенью возможности. Т.е. в некоторых случаях можно сказать, что одно событие произойдет практически наверняка, другое практически никогда.

В отношении друг друга события также имеют особенности, т.е. в одном случае событие А может произойти совместно с событием В, в другом – нет.

<u>Определение.</u> События называются **несовместными**, если появление одного из них исключает появление других.

Классическим примером несовместных событий является результат подбрасывания монеты – выпадение лицевой стороны монеты исключает выпадение обратной стороны (в одном и том же опыте).

<u>Определение.</u> **Полной группой событий** называется совокупность всех возможных результатов опыта.

<u>Определение.</u> **Достоверным событием** называется событие, которое наверняка произойдет в результате опыта. Событие называется **невозможным**, если оно никогда не произойдет в результате опыта.

Например, если из коробки, содержащей только красные и зеленые шары, наугад вынимают один шар, то появление среди вынутых шаров белого – невозможное событие. Появление красного и появление зеленого шаров образуют полную группу событий.

<u>Определение.</u> События называются **равновозможными**, если нет оснований считать, что одно из них появится в результате опыта чаще, чем другие.

<u>Определение.</u> **Вероятностью** события А называется математическая оценка возможности появления этого события в результате опыта.

Исход опыта является благоприятствующим событию А, если появление в результате опыта этого исхода влечет за собой появление события А.

<u>Определение.</u> Вероятностью события А или классической вероятностью называется отношение числа, благоприятствующих событию А исходов опыта к общему числу попарно несовместных исходов опыта, образующих полную группу событий.

$$P(A) = \frac{m}{n}$$

Очевидно, что вероятность достоверного события равна единице, а вероятность невозможного – равна нулю. Таким образом, значение вероятности любого события – есть положительное число, заключенное между нулем и единицей.

$$0 \le P(A) \le 1$$

<u>Пример.</u> В коробке находится 10 шаров. 3 из них красные, 2 – зеленые, остальные белые. Найти вероятность того, что вынутый наугад шар будет красным, зеленым или белым.

Появление красного, зеленого и белого шаров составляют полную группу событий. Обозначим появление красного шара – событие А, появление зеленого – событие В, появление белого – событие С.

Тогда в соответствием с записанными выше формулами получаем:

$$P(A) = \frac{3}{10}; \quad P(B) = \frac{2}{10}; \quad P(C) = \frac{5}{10}.$$

<u>**Определение.**</u> **Относительной частотой** события А называется отношение числа опытов, в результате которых произошло событие А к общему числу опытов.

Отличие относительной частоты от вероятности заключается в том, что вероятность вычисляется без непосредственного произведения опытов, а относительная частота – после опыта.

Так в рассмотренном выше примере, если из коробки наугад извлечено 5 шаров и 2 из них оказались красными, то относительная частота появления красного шара равна:

$$W(A) = \frac{2}{5}.$$

Как видно, эта величина не совпадает с найденной вероятностью.

При достаточно большом числе произведенных опытов относительная частота изменяется мало, колеблясь около одного числа. Это число может быть принято за вероятность события. Его называют статистической вероятностью.

Классическое определение вероятности неприменимо к испытаниям с бесконечным числом исходов. Чтобы преодолеть этот недостаток вводится понятие **геометрической вероятности**, т.е. вероятности попадания точки в какой – либо отрезок или часть плоскости (пространства).

Так если на отрезке длиной L выделен отрезок длины l, то вероятность попадания наугад взятой точки в отрезок l равна отношению l/L.

Операции над событиями.

Определение. События А и В называются **равными**, если осуществление события А влечет за собой осуществление события В и наоборот.

Определение. **Объединением** или **суммой** событий A_k называется событие А, которое означает появление **хотя бы одного** из событий A_k.

$$A = \bigcup_k A_k$$

Определение. **Пересечением** или **произведением** событий A_k называется событие А, которое заключается в осуществлении **всех** событий A_k.

$$A = \bigcap_k A_k$$

Определение. **Разностью** событий А и В называется событие С, которое означает, что происходит событие А, но не происходит событие В.

$$C = A \setminus B$$

Определение. **Противоположным** к событию А называется событие $\overline{A}$, означающее, что событие А **не происходит.**

Определение. **Элементарными исходами** опыта называются такие результаты опыта, которые взаимно исключают друг друга и в результате опыта происходит одно из этих событий, также каково бы ни было событие

А, по наступившему элементарному исходу можно судить о том, происходит или не происходит это событие.

Совокупность всех элементарных исходов опыта называется **пространством элементарных событий**.

Лекция 2.

Теорема сложения вероятностей. Противоположные события. Условные вероятности. Теорема умножения вероятностей. Независимые события. Вероятность появления хотя бы одного события.

Теорема (сложения вероятностей). *Вероятность суммы двух несовместных событий равна сумме вероятностей этих событий.*

$$P(A+B)=P(A)+P(B).$$

Следствие 1: *Если события $A_1, A_2, ..., A_n$ образуют полную группу несовместных событий, то сумма их вероятностей равна единице.*

$$\sum_{i=1}^{n} P(A_i)=1.$$

Определение. Противоположными называются два несовместных события, образующие полную группу.

Теорема. *Вероятность появления хотя бы одного из двух совместных событий равна сумме вероятностей этих событий без вероятности их совместного появления.*

$$P(A+B)=P(A)+P(B)-P(AB).$$

Следствие 2: *Сумма вероятностей противоположных событий равна единице.*

$$P(A)+P(\overline{A})=1.$$

Определение. Событие А называется **независимым** от события В, если вероятность события А не зависит от того, произошло событие В или нет. Событие А называется **зависимым** от события В, если вероятность события А меняется в зависимости от того, произошло событие В или нет.

<u>Определение.</u> Вероятность события В, вычисленная при условии, что имело место событие А, называется **условной вероятностью** события В.

$$P_A(B) = P(B/A) = P(AB)/P(A).$$

<u>Теорема.</u> (Умножения вероятностей) *Вероятность произведения двух событий (совместного появления этих событий) равна произведению вероятности одного из них на условную вероятность другого, вычисленную при условии, что первое событие уже наступило.*

$$P(AB) = P(A)P(B/A) = P(A)P_A(B)$$

Также можно записать: $P(AB) = P(A)P(B/A) = P(B)P(A/B) = P(B)P_B(A)$.

Если события независимые, то $P(B/A) = P(B)$, и теорема умножения вероятностей принимает вид:

$$P(AB) = P(A)P(B).$$

В случае произведения нескольких зависимых событий вероятность равна произведению одного из них на условные вероятности всех остальных при условии, что вероятность каждого последующего вычисляется в предположении, что все предыдущие события уже совершились.

$$P(A_1A_2...A_n) = P(A_1)P(A_2/A_1)P(A_3/A_1A_2)...P(A_n/A_1A_2...A_{n-1})$$

Из теоремы произведения вероятностей можно сделать вывод о вероятности появления хотя бы одного события.

Если в результате испытания может появиться n событий, независимых в совокупности, то вероятность появления хотя бы одного из них равна

$$P(A) = 1 - q_1q_2...q_n$$

Здесь событие А обозначает наступление хотя бы одного из событий A_i, а q_i – вероятность противоположных событий $\overline{A}_1, \overline{A}_2, ..., \overline{A}_n$.

<u>Пример.</u> Чему равна вероятность того, что при бросании трех игральных костей 6 очков появится хотя бы на одной из костей?

Вероятность выпадения 6 очков при одном броске кости равна $\frac{1}{6}$. Вероятность того, что не выпадет 6 очков – $\frac{5}{6}$. Вероятность того, что при броске трех костей не выпадет ни разу 6 очков, равна $p = \left(\frac{5}{6}\right)^3 = \frac{125}{216}$.

Тогда вероятность того, что хотя бы один раз выпадет 6 очков равна $1 - \frac{125}{216} = \frac{91}{216}$.

Пример. Два стрелка стреляют по мишени. Вероятность попадания в мишень при одном выстреле для первого стрелка равна 0,7, а для второго – 0,8. Найти вероятность того, что при одном залпе в мишень попадает только один из стрелков.

Обозначим попадание в цель первым стрелком – событие A, вторым – событие B, промах первого стрелка – событие $\overline{A}$, промах второго – событие $\overline{A}$. $P(A) = 0{,}7;\quad P(\overline{A}) = 0{,}3;\quad P(B) = 0{,}8;\quad P(\overline{B}) = 0{,}2.$

Вероятность того, что первый стрелок попадет в мишень, а второй – нет равна $P(A)P(\overline{B}) = 0{,}7 \cdot 0{,}2 = 0{,}14$

Вероятность того, что второй стрелок попадет в цель, а первый – нет равна $P(\overline{A})P(B) = 0{,}3 \cdot 0{,}8 = 0{,}24$

Тогда вероятность попадания в цель только одним стрелком равна

$$P = 0{,}14 + 0{,}24 = 0{,}38$$

Тот же результат можно получить другим способом – находим вероятности того, что оба стрелка попали в цель и оба промахнулись. Эти вероятности соответственно равны:

$$P(A)P(B) = 0{,}7 \cdot 0{,}8 = 0{,}56;\quad P(\overline{A})P(\overline{B}) = 0{,}3 \cdot 0{,}2 = 0{,}06.$$

Тогда вероятность того, что в цель попадет только один стрелок равна:

$$P = 1 - 0{,}56 - 0{,}06 = 0{,}38.$$

$$P = 3 \cdot 0{,}2 \cdot 0{,}8 \cdot 0{,}8 = 0{,}384$$

Лекция 3.

Формула полной вероятности и формула Байеса.

Формула полной вероятности.

Пусть некоторое событие А может произойти вместе с одним из несовместных событий $H_1, H_2, ..., H_n$, составляющих полную группу событий. Пусть известны вероятности этих событий $P(H_1), P(H_2), ..., P(H_n)$ и условные вероятности наступления события А при наступлении события H_i $P(A/H_1), P(A/H_2), ..., P(A/H_n)$.

Теорема. *Вероятность события А, которое может произойти вместе с одним из событий* $H_1, H_2, ..., H_n$, *равна сумме парных произведений вероятностей каждого из этих событий на соответствующие им условные вероятности наступления события А.*

$$P(A) = \sum_{i=1}^{n} P(H_i)P(A/H_i)$$

Доказательство.

Т.к. события $H_1, H_2, ..., H_n$ образуют полную группу событий, то событие А можно представить в виде следующей суммы:

$$A = AH_1 + AH_2 + ... + AH_n = \sum_{i=1}^{n} AH_i$$

Т.к. события $H_1, H_2, ..., H_n$ несовместны, то и события AH_i тоже несовместны. Тогда можно применить теорему о сложении вероятностей несовместных событий: $P(A) = \sum_{i=1}^{n} P(AH_i)$.

При этом $P(AH_i) = P(H_i)P(A/H_i)$.

Окончательно получаем: $P(A) = \sum_{i=1}^{n} P(H_i)P(A/H_i)$.

Формула Байеса (формула гипотез).

Пусть имеется полная группа несовместных гипотез $H_1, H_2, ..., H_n$ с известными вероятностями их наступления $P(H_1), P(H_2), ..., P(H_n)$. Пусть в результате опыта наступило событие А, условные вероятности которого по каждой из гипотез известны, т.е. известны вероятности $P(A/H_1), P(A/H_2), ..., P(A/H_n)$.

Требуется определить какие вероятности имеют гипотезы $H_1, H_2, ..., H_n$ относительно события А, т.е. условные вероятности $P(H_i/A)$.

Теорема. *Вероятность гипотезы после испытания равна произведению вероятности гипотезы до испытания на соответствующую ей условную вероятность события, которое произошло при испытании, деленному на полную вероятность этого события.*

$$P(H_i/A) = \frac{P(H_i)P(A/H_i)}{\sum_{i=1}^{n} P(H_i)P(A/H_i)}$$

Эта формула называется **формулой Байеса**.

Доказательство.

По Теореме умножения вероятностей

получаем: $P(A)P(H_i/A) = P(H_i)P(A/H_i)$

Тогда если $P(A) \neq 0$, $\quad P(H_i/A) = \dfrac{P(H_i)P(A/H_i)}{P(A)}$.

Для нахождения вероятности P(A) используем формулу полной вероятности.

$$P(H_i/A) = \frac{P(H_i)P(A/H_i)}{\sum_{i=1}^{n} P(H_i)P(A/H_i)}.$$

Лекция 4. Схема и формула Бернулли. Приближение Пуассона для схемы Бернулли. Функция Лапласа.

Если производится некоторое количество испытаний, в результате которых может произойти или не произойти событие А, и вероятность появления этого события в каждом из испытаний не зависит от результатов остальных испытаний, то такие испытания называются **независимыми относительно события А**.

Допустим, что событие А наступает в каждом испытании с вероятностью *P(A)=p*. Определим вероятность $P_{m,n}$ того, что в результате *n* испытаний событие А наступило ровно *m* раз.

Пусть в результате каждого из *n* независимых испытаний, проведенных в одинаковых условиях, событие А наступает с вероятностью *P(A) = p*, а противоположное ему событие $\overline{A}$ с вероятностью $P(\overline{A}) = 1 - p$.

Обозначим A_i – наступление события А в испытании с номером *i*. Т.к. условия проведения опытов одинаковые, то эти вероятности равны.

Если в результате *n* опытов событие А наступает ровно *m* раз, то остальные *n-m* раз это событие не наступает. Событие А может появиться *m* раз в *n* испытаниях в различных комбинациях, число которых равно количеству сочетаний из *n* элементов по *m*. Это количество сочетаний находится по формуле:

$$C_n^m = \frac{n!}{m!(n-m)!}$$

Вероятность каждой комбинации равна произведению вероятностей:

$$p^m (1-p)^{n-m}$$

Применяя теорему сложения вероятностей несовместных событий, получаем **формулу Бернулли:**

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

Формула Бернулли важна тем, что справедлива для любого количества независимых испытаний, т.е. того самого случая, в котором наиболее четко проявляются законы теории вероятностей.

Если число испытаний *n* велико, то применение формулы Бернулли приводит к очень громоздким вычислениям. В таких случаях пользуются приближенными формулами, основанными на теоремах Лапласа и Пуассона.

Локальная теорема Лапласа. Вероятность $P_n(m)$ того, что в n испытаниях, в каждом из которых вероятность появления события равна *p*, событие наступит ровно m раз приближенно равна:

$$P_n(m)\approx\frac{\varphi(x)}{\sqrt{npq}},\quad \text{где}\quad x=\frac{m-np}{\sqrt{npq}},\qquad \varphi(x)=\frac{1}{\sqrt{2\pi}}e^{-\frac{x^2}{2}}.$$

Значение функции φ(x) находится по таблице.

Так как $\varphi(x)$ – четная функция, т.е. $\varphi(-x)=\varphi(x)$, то таблица составлена только для положительных x. Для всех $x > 5$ считают, что $\varphi(x)=0$.

Формула Пуассона. Если вероятность p наступления события постоянна и мала, а число испытаний n велико, то:

$$P_n(m)\approx\frac{\lambda^m\cdot e^{-\lambda}}{m!},\quad \text{где}\quad \lambda=n\,p.$$

Число m_0 наступления события A в n независимых испытаниях называется наивероятнейшим, если вероятность наступления события данное число раз наибольшая по сравнению с вероятностями других исходов.

Наивероятнейшее число m_0 определяется из двойного неравенства

$$np-q\le m_0\le np+p,$$

где p – вероятность появления события A, $q = 1 - p$.

При этом, если $n\,p + p$ и $n\,p - q$ – целые числа, то наивероятнейших чисел два.

Задача 1. Работница обслуживает 800 веретен. Вероятность обрыва пряжи на каждом из них за промежуток времени t равна 0,005. Найдите наивероятнейшее число обрывов и его вероятность.

Решение. Имеем $n = 800$, $p = 0{,}005$, найдем $q = 1 - p = 0{,}095$.

Запишем неравенство для нахождения наивероятнейшего числа:

$$800\cdot 0{,}005-0{,}995\le m_0\le 800\cdot 0{,}005.$$

Откуда наивероятнейшее число обрыва пряжи $m_0 = 4$.

Вероятность $P_{800}(4) = 0{,}1945$.

Задача 2. Новые приборы при изготовлении подвергаются всесторонним испытаниям на прочность, вибрацию и др. видов перегрузок.

Предположим, что вероятность Р опасной перегрузки для прибора при единичном испытании равна 0,4. Экспериментатор решил провести с прибором 3 опыта. При этом известно, что вероятность поломки приборов при однократной перегрузке 0,2; при 2 – кратной – 0,5; при 3-кратной – 0,8. Имеется опасность, что прибор после проведения этих опытов может поломаться, что нежелательно. Определить вероятность поломки при 3 опытах.

Решение.

H_1 – однократная перегрузка прибора,

H_2 – 2-кратная перегрузка прибора,

H_3 – 3-кратная перегрузка прибора,

A – поломка прибора.

По формуле Бернулли находим:

$$P(H_1) = P(1) = C_3^1\, 0{,}4^1 \cdot 0{,}6^2 = 0{,}432,$$

$$P(H_2) = P(2) = C_3^{21}\, 0{,}4^2 \cdot 0{,}6^1 = 0{,}288,$$

$$P(H_3) = P(3) = C_3^3\, 0{,}4^3 \cdot 0{,}6^0 = 0{,}061,$$

Далее имеем: $P_{H_1}(A) = 0{,}2$, $P_{H_2}(A) = 0{,}5$ $P_{H_3}(A) = 0{,}8.$

По формуле полной вероятности вычислим вероятность поломки прибора при 3 испытания

$$P(A) = 0{,}432 \cdot 0{,}2 + 0{,}288 \cdot 0{,}5 + 0{,}061 \cdot 0{,}8 = 0{,}2816,$$

т.е. риск экспериментатора весьма значителен.

Лекция 5.

Случайные величины. Закон распределения и функция распределения дискретной случайной величины. Биномиальное распределение и распределение Пуассона.

Случайные величины.

Выше рассматривались случайные события, являющиеся качественной характеристикой случайного результата опыта. Для получения количественной характеристики вводится понятие случайной величины.

Определение. **Случайной величиной** называется величина, которая в результате опыта может принимать то или иное значение, причем заранее не известно какое именно.

Случайные величины можно разделить на две категории.

Определение. **Дискретной случайной величиной** называется такая величина, которая в результате опыта может принимать определенные значения с определенной вероятностью, образующие счетное множество (множество, элементы которого могут быть занумерованы).

Это множество может быть как конечным, так и бесконечным.

Например, количество выстрелов до первого попадания в цель является дискретной случайной величиной, т.к. эта величина может принимать и бесконечное, хотя и счетное количество значений.

Определение. **Непрерывной случайной величиной** называется такая величина, которая может принимать любые значения из некоторого конечного или бесконечного промежутка.

Очевидно, что число возможных значений непрерывной случайной величины бесконечно.

Для задания случайной величины недостаточно просто указать ее значение, необходимо также указать вероятность этого значения.

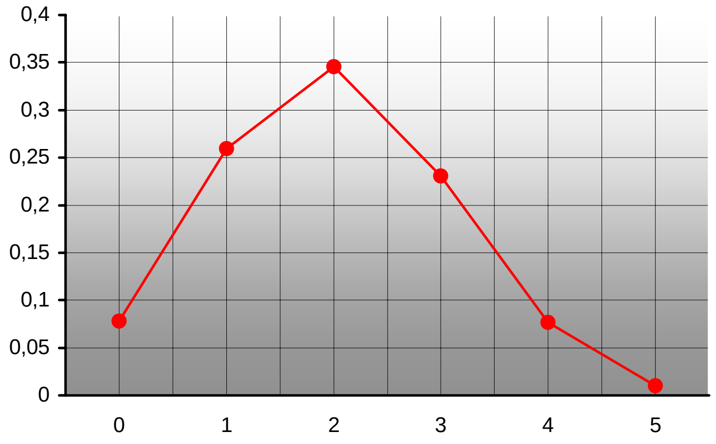
Закон распределения дискретной случайной величины.

Определение. Соотношение между возможными значениями случайной величины и их вероятностями называется **законом распределения дискретной** случайной величины.

Закон распределения может быть задан аналитически, в виде таблицы или графически.

Таблица соответствия значений случайной величины и их вероятностей называется **рядом распределения**.

Графическое представление этой таблицы называется **многоугольником распределения.** При этом сумма все ординат



многоугольника распределения представляет собой вероятность всех возможных значений случайной величины, а, следовательно, равна единице.

При построении многоугольника распределения надо помнить, что соединение полученных точек носит условный характер. В промежутках между значениями случайной величины вероятность не принимает никакого значения. Точки соединены только для наглядности.

<u>Биноминальное распределение.</u>

Если производится n независимых испытаний, в каждом из которых событие А может появиться с одинаковой вероятностью p в каждом из испытаний, то вероятность того, что событие не появится, равна $q = 1 - p$.

Примем число появлений события в каждом из испытаний за некоторую случайную величину Х.

Чтобы найти закон распределения этой случайной величины, необходимо определить значения этой величины и их вероятности.

Значения найти достаточно просто. Очевидно, что в результате n испытаний событие может не появиться вовсе, появиться один раз, два раза, три и т.д. до n раз.

Вероятность каждого значения этой случайной величины можно найти по формуле Бернулли.

$$P_n(k) = C_n^k p^k q^{n-k}, \qquad k = 0,1,2,...$$

Эта формула аналитически выражает искомый закон распределения. Этот закон распределения называется **биноминальным**.

Пример. В партии 10% нестандартных деталей. Наугад отобраны 4 детали. Написать биноминальный закон распределения дискретной случайной величины X – числа нестандартных деталей среди четырех отобранных и построить многоугольник полученного распределения.

Вероятность появления нестандартной детали в каждом случае равна 0,1.

Найдем вероятности того, что среди отобранных деталей:

1) Вообще нет нестандартных. $P_4(0) = \frac{4!}{0! \cdot 4!} 0{,}1^0 \cdot 0{,}9^4 = 0{,}6561$

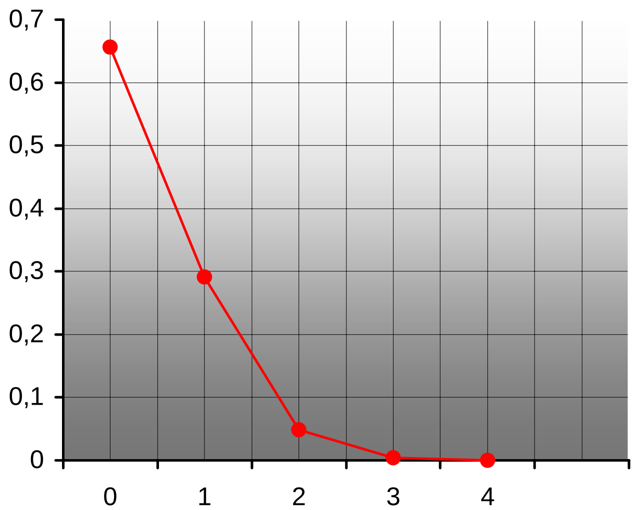
2) Одна нестандартная. $P_4(1) = \frac{4!}{1! \cdot 3!} 0{,}1^1 \cdot 0{,}9^3 = 0{,}2916$

3) Две нестандартные детали. $P_4(2) = \frac{4!}{2! \cdot 2!} 0{,}1^2 \cdot 0{,}9^2 = 0{,}0486$

4) Три нестандартные детали. $P_4(3) = \frac{4!}{3! \cdot 1!} 0{,}1^3 \cdot 0{,}9^1 = 0{,}0036$

5) Четыре нестандартных детали. $P_4(4) = \frac{4!}{4! \cdot 0!} 0{,}1^4 \cdot 0{,}9^0 = 0{,}0001$

Построим многоугольник распределения.



<u>Распределение Пуассона.</u>

(Симеон Дени Пуассон (1781 – 1840) – французский математик)

Пусть производится n независимых испытаний, в которых появление события А имеет вероятность p. Если число испытаний n достаточно велико, а вероятность появления события А в каждом испытании мала ($p \le 0{,}1$), то для нахождения вероятности появления события А k раз находится следующим образом.

Сделаем важное допущение – произведение np сохраняет постоянное значение: $np = \lambda$

Практически это допущение означает, что среднее число появления события в различных сериях испытаний (при разном n) остается неизменным.

По формуле Бернулли получаем:

$$P_n(k) = \frac{n(n-1)(n-2)...(n-(k-1))}{k!} p^k (1-p)^{n-k}$$

$$P_n(k) = \frac{n(n-1)(n-2)...(n-(k-1))}{k!} \left(\frac{\lambda}{n}\right)^k \left(1-\frac{\lambda}{n}\right)^{n-k}$$

Найдем предел этой вероятности при $n \to \infty$.

$$= \frac{\lambda^k}{k!} \lim_{n\to\infty} \left[\left(1-\frac{1}{n}\right)\left(1-\frac{2}{n}\right)...\left(1-\frac{k-1}{n}\right)\left(1-\frac{\lambda}{n}\right)^{n-k}\right] = \frac{\lambda^k}{k!} \lim_{n\to\infty}\left(1-\frac{\lambda}{n}\right)^n \lim_{n\to\infty}\left(1-\frac{\lambda}{n}\right)^{-k} = \frac{\lambda^k}{k!} e^{-\lambda}$$

$$P_n(k) \cong \lim_{n\to\infty} \frac{n(n-1)(n-2)...(n-(k-1))}{k!} \frac{\lambda^k}{n^k} \left(1-\frac{\lambda}{n}\right)^{n-k} =$$

Получаем формулу **распределения Пуассона:**

$$P_n(k) = \frac{\lambda^k e^{-\lambda}}{k!}$$

Если известны числа λ и k, то значения вероятности можно найти по соответствующим таблицам распределения Пуассона.

Лекция 6.

Основные числовые характеристики дискретных и непрерывных случайных величин: математическое ожидание, дисперсия и среднее квадратическое отклонение. Их свойства и примеры.

Числовые характеристики дискретных случайных величин.

Закон распределения полностью характеризует случайную величину. Однако когда невозможно найти закон распределения, или этого не требуется, можно ограничиться нахождением значений, называемых числовыми характеристиками случайной величины. Эти величины определяют некоторое среднее значение, вокруг которого группируются значения случайной величины, и степень их разбросанности вокруг этого среднего значения.

Определение. **Математическим ожиданием** дискретной случайной величины называется сумма произведений всех возможных значений случайной величины на их вероятности.

$$m_x = M(X) = x_1 p_1 + x_2 p_2 + \ldots + x_n p_n = \sum_{i=1}^{n} x_i p_i$$

С точки зрения вероятности можно сказать, что математическое ожидание приближенно равно среднему арифметическому наблюдаемых значений случайной величины.

Свойства математического ожидания.

1) Математическое ожидание постоянной величины равно самой постоянной.

$$M(C) = C$$

2) Постоянный множитель можно выносить за знак математического ожидания.

$$M(Cx) = CM(x)$$

3) Математическое ожидание произведения двух независимых случайных величин равно произведению их математических ожиданий.

$$M(XY) = M(X)M(Y)$$

Это свойство справедливо для произвольного числа случайных величин.

4) Математическое ожидание суммы двух случайных величин равно сумме математических ожиданий слагаемых.

$$M(X+Y) = M(X) + M(Y)$$

Это свойство также справедливо для произвольного числа случайных величин.

Пусть производится n независимых испытаний, вероятность появления события А в которых равна p.

Теорема. *Математическое ожидание M(X) числа появления события A в n независимых испытаниях равно произведению числа испытаний на вероятность появления события в каждом испытании.*

$$M(X) = np$$

Однако математическое ожидание не может полностью характеризовать случайный процесс. Кроме математического ожидания надо ввести величину, которая характеризует отклонение значений случайной величины от математического ожидания.

Это отклонение равно разности между случайной величиной и ее математическим ожиданием. При этом математическое ожидание отклонения равно нулю. Это объясняется тем, что одни возможные отклонения положительны, другие отрицательны, и в результате их взаимного погашения получается ноль.

Определение. Дисперсией (рассеиванием) дискретной случайной величины называется математическое ожидание квадрата отклонения случайной величины от ее математического ожидания.

$$D(X) = M\left[X - M(X)\right]^2$$

Однако, на практике подобный способ вычисления дисперсии неудобен, т.к. приводит при большом количестве значений случайной величины к громоздким вычислениям.

Поэтому применяется другой способ

Вычисление дисперсии.

Теорема. *Дисперсия равна разности между математическим ожиданием квадрата случайной величины X и квадратом ее математического ожидания.*

$$D(X) = M(X^2) - [M(X)]^2$$

Доказательство. С учетом того, что математическое ожидание M(X) и квадрат математического ожидания $M^2(X)$ – величины постоянные, можно записать:

$$D(X) = M[X - M(X)]^2 = M[X^2 - 2XM(X) + M^2(X)] =$$

$$= M(X^2) - 2M(X)M(X) + M^2(X) = M(X^2) - M^2(X).$$

$$D(X) = M(X^2) - [M(X)]^2$$

Свойства дисперсии.

1) Дисперсия постоянной величины равна нулю.

$$D(C) = 0$$

2) Постоянный множитель можно выносить за знак дисперсии, возводя его в квадрат.

$$D(CX) = C^2 D(X)$$

3) Дисперсия суммы двух независимых случайных величин равна сумме дисперсий этих величин.

$$D(X + Y) = D(X) + D(Y)$$

4) Дисперсия разности двух независимых случайных величин равна сумме дисперсий этих величин.

$$D(X - Y) = D(X) + D(Y)$$

Справедливость этого равенства вытекает из свойства 2.

Теорема. *Дисперсия числа появления события А в n независимых испытаний, в каждом из которых вероятность р появления события постоянна, равна произведению числа испытаний на вероятности появления и непоявления события в каждом испытании.*

$$D(X) = npq$$

Среднее квадратическое отклонение.

Определение. Средним квадратическим отклонением случайной величины Х называется квадратный корень из дисперсии.

$$\sigma(X) = \sqrt{D(X)}$$

Теорема. *Среднее квадратичное отклонение суммы конечного числа взаимно независимых случайных величин равно квадратному корню из суммы квадратов средних квадратических отклонений этих величин.*

$$\sigma(X_1 + X_2 + ... + X_n) = \sqrt{\sigma^2(X_1) + \sigma^2(X_2) + ... + \sigma^2(X_n)}$$

Пример. Испытывается устройство, состоящее из четырех независимо работающих приборов. Вероятности отказа каждого из приборов равны соответственно $p_1=0,3;$ $p_2=0,4;$ $p_3=0,5;$ $p_4=0,6.$ Найти математическое ожидание и дисперсию числа отказавших приборов.

Принимая за случайную величину число отказавших приборов, видим что эта случайная величина может принимать значения 0, 1, 2, 3 или 4.

Для составления закона распределения этой случайной величины необходимо определить соответствующие вероятности. Примем $q_i = 1 - p_i$.

1) Не отказал ни один прибор.

$$p(0) = q_1q_2q_3q_4 = 0,7 \cdot 0,4 \cdot 0,5 \cdot 0,4 = 0,084.$$

2) Отказал один из приборов.

$$p(1) = p_1q_2q_3q_4 + q_1p_2q_3q_4 + q_1q_2p_3q_4 + q_1q_2q_3p_4 = 0,302.$$

3) Отказали два прибора.

$$p(2) = p_1p_2q_3q_4 + p_1q_2p_3q_4 + p_1q_2q_3p_4 + q_1p_2p_3q_4 + q_1p_2q_3p_4 + q_1q_2p_3p_4 = 0,38.$$

4) Отказали три прибора.

$$p(3) = p_1 p_2 p_3 q_4 + p_1 p_2 q_3 p_4 + p_1 q_2 p_3 p_4 + q_1 p_2 p_3 p_4 = 0{,}198.$$

5) Отказали все приборы. $p(4) = p_1 p_2 p_3 p_4 = 0{,}036.$

Получаем закон распределения:

x	0	1	2	3	4
x^2	0	1	4	9	16
p	0,084	0,302	0,38	0,198	0,036

Математическое ожидание:

$$M(X) = 0{,}302 + 2 \cdot 0{,}38 + 3 \cdot 0{,}198 + 4 \cdot 0{,}036 = 1{,}8.$$

$$M(X^2) = 0{,}302 + 4 \cdot 0{,}38 + 9 \cdot 0{,}198 + 16 \cdot 0{,}036 = 4{,}18.$$

Дисперсия: $D(X) = M(X^2) - \left[M(X)\right]^2 = 4{,}18 - 3{,}24 = 0{,}94.$

Лекция 7.

Функция распределения и плотность распределения непрерывной случайной величины, их взаимосвязь и свойства. Равномерное распределение вероятностей.

Функция распределения.

Во всех рассмотренных выше случаях случайная величина определялась путем задания значений самой величины и вероятностей этих значений.

Однако, такой метод применим далеко не всегда. Например, в случае непрерывной случайной величины, ее значения могут заполнять некоторый произвольный интервал. Очевидно, что в этом случае задать все значения случайной величины просто нереально.

Даже в случае, когда это сделать можно, зачастую задача решается чрезвычайно сложно. Рассмотренный только что пример даже при относительно простом условии (приборов только четыре) приводит к достаточно неудобным вычислениям, а если в задаче будет несколько сотен приборов?

Поэтому встает задача по возможности отказаться от индивидуального подхода к каждой задаче и найти по возможности наиболее общий способ задания любых типов случайных величин.

Пусть x – действительное число. Вероятность события, состоящего в том, что X примет значение, меньшее x, т.е. $X < x$, обозначим через F(x).

<u>**Определение.**</u> **Функцией распределения** называют функцию F(x), определяющую вероятность того, что случайная величина X в результате испытания примет значение, меньшее x.

$$F(x) = P(X < x)$$

Функцию распределения также называют **интегральной функцией**. Функция распределения существует как для непрерывных, так и для дискретных случайных величин. Она полностью характеризует случайную величину и является одной из форм закона распределения.

Для дискретной случайной величины функция распределения имеет вид:

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

Знак неравенства под знаком суммы показывает, что суммирование распространяется на те возможные значения случайной величины, которые меньше аргумента x.

Функция распределения дискретной случайной величины X разрывна и возрастает скачками при переходе через каждое значение x_i.

<u>Свойства функции распределения..</u>

1) значения функции распределения принадлежат отрезку [0, 1].

$$0 \le F(x) \le 1$$

2) F(x) – неубывающая функция.

$$F(x_2) \ge F(x_1) \text{ при } x_2 \ge x_1$$

3) Вероятность того, что случайная величина примет значение, заключенное в интервале (a, b) , равна приращению функции распределения на этом интервале.

$$P(a \le X < b) = F(b) - F(a)$$

4) На минус бесконечности функция распределения равна нулю, на плюс бесконечности функция распределения равна единице.

5) Вероятность того, что непрерывная случайная величина X примет одно определенное значение, равна нулю.

Таким образом, не имеет смысла говорить о каком – либо конкретном значении случайной величины. Интерес представляет только вероятность попадания случайной величины, в какой – либо интервал, что соответствует большинству практических задач.

Плотность распределения.

Функция распределения полностью характеризует случайную величину, однако, имеет один недостаток. По функции распределения трудно судить о характере распределения случайной величины в небольшой окрестности той или иной точки числовой оси.

Определение. **Плотностью распределения** вероятностей непрерывной случайной величины X называется функция *f(x)* – первая производная от функции распределения F(x).

$$f(x) = F'(x).$$

Плотность распределения также называют **дифференциальной функцией**. Для описания дискретной случайной величины плотность распределения неприемлема.

Смысл плотности распределения состоит в том, что она показывает как часто появляется случайная величина X в некоторой окрестности точки *x* при повторении опытов.

После введения функций распределения и плотности распределения можно дать следующее определение непрерывной случайной величины.

<u>**Определение.**</u> Случайная величина X называется **непрерывной**, если ее функция распределения F(x) непрерывна на всей оси OX, а плотность распределения f(x) существует везде, за исключением (может быть конечного числа точек).

Зная плотность распределения, можно вычислить вероятность того, что некоторая случайная величина X примет значение, принадлежащее заданному интервалу.

<u>**Теорема.**</u> *Вероятность того, что непрерывная случайная величина X примет значение, принадлежащее интервалу (a, b), равна определенному интегралу от плотности распределения, взятому в пределах от a до b.*

$$P(a < X < b) = \int_a^b f(x)dx$$

Доказательство этой теоремы основано на определении плотности распределения и третьем свойстве функции распределения, записанном выше.

Геометрически это означает, что вероятность того, что непрерывная случайная величина примет значение, принадлежащее интервалу (a, b), равна площади криволинейной трапеции, ограниченной осью OX, кривой распределения f(x) и прямыми $x=a$ и $x=b$.

Функция распределения может быть легко найдена, если известна плотность распределения, по формуле:

$$F(x) = \int_{-\infty}^{x} f(x)dx$$

<u>Свойства плотности распределения.</u>

1) Плотность распределения – неотрицательная функция.

$$f(x) \geq 0$$

2) Несобственный интеграл от плотности распределения в пределах от - ∞ до ∞ равен единице. $\int_{-\infty}^{\infty} f(x)dx = 1.$

Лекция 8.

Основные числовые характеристики непрерывных случайных величин: математическое ожидание, дисперсия и среднее квадратическое отклонение. Их свойства и примеры.

Числовые характеристики непрерывных случайных величин.

Пусть непрерывная случайная величина X задана функцией распределения *f(x)*. Допустим, что все возможные значения случайной величины принадлежат отрезку [*a,b*].

Определение. Математическим ожиданием непрерывной случайной величины X, возможные значения которой принадлежат отрезку [a,b], называется определенный интеграл

$$M(X) = \int_{a}^{b} xf(x)dx$$

Если возможные значения случайной величины рассматриваются на всей числовой оси, то математическое ожидание находится по формуле:

$$M(X) = \int_{-\infty}^{\infty} xf(x)dx$$

При этом, конечно, предполагается, что несобственный интеграл сходится.

Определение. Дисперсией непрерывной случайной величины называется математическое ожидание квадрата ее отклонения.

$$D(X) = \int_{-\infty}^{\infty} [x - M(X)]^2 f(x)dx$$

По аналогии с дисперсией дискретной случайной величины, для практического вычисления дисперсии используется формула:

$$D(X) = \int_{-\infty}^{\infty} x^2 f(x)dx - [M(X)]^2$$

<u>Определение.</u> Средним квадратичным отклонением называется квадратный корень из дисперсии.

$$\sigma(X) = \sqrt{D(X)}$$

<u>Определение.</u> Модой M_0 дискретной случайной величины называется ее наиболее вероятное значение. Для непрерывной случайной величины мода – такое значение случайной величины, при которой плотность распределения имеет максимум. $f(M_0) = \max$.

Если многоугольник распределения для дискретной случайной величины или кривая распределения для непрерывной случайной величины имеет два или несколько максимумов, то такое распределение называется **двухмодальным** или **многомодальным**.

Если распределение имеет минимум, но не имеет максимума, то оно называется **антимодальным**.

<u>Определение.</u> Медианой M_D случайной величины X называется такое ее значение, относительно которого равновероятно получение большего или меньшего значения случайной величины.

$$P(X < M_D) = P(X > M_D)$$

Геометрически медиана – абсцисса точки, в которой площадь, ограниченная кривой распределения делится пополам.

<u>Определение.</u> Начальным моментом порядка k случайной величины X называется математическое ожидание величины X^k.

$$\alpha_k = M[X^k].$$

Для дискретной случайной величины:

$$\alpha_k = \sum_{i=1}^{n} x_i^k p_i \ .$$

Для непрерывной случайной величины:

$$\alpha_k = \int_{-\infty}^{\infty} x^k f(x)dx .$$

Начальный момент первого порядка равен математическому ожиданию.

<u>Определение.</u> **Центральным моментом** порядка k случайной величины X называется математическое ожидание величины $(X - m_x)^k$.

$$\mu_k = M[(X - m_x)^k]$$

Для дискретной случайной величины:

$$\mu_k = \sum_{i=1}^{n} (x_i - m_x)^k p_i .$$

Для непрерывной случайной величины:

$$\mu_k = \int_{-\infty}^{\infty} (x - m_x)^k f(x)dx .$$

Центральный момент первого порядка всегда равен нулю, а центральный момент второго порядка равен дисперсии. Центральный момент третьего порядка характеризует асимметрию распределения.

<u>Определение.</u> Отношение центрального момента третьего порядка к среднему квадратическому отклонению в третьей степени называется **коэффициентом асимметрии**.

$$a_x = \frac{\mu_3}{\sigma_x^3}$$

<u>Определение.</u> Для характеристики островершинности и плосковершинности распределения используется величина, называемая **эксцессом**.

$$C_x = \frac{\mu_4}{\sigma_x^4} - 3$$

Кроме рассмотренных величин используются также так называемые абсолютные моменты:

Абсолютный начальный момент:

$$\beta_k = M[|X|^k] .$$

Абсолютный центральный момент:

$$\nu_k = M[|X - m_x|^k].$$

Абсолютный центральный момент первого порядка называется **средним арифметическим отклонением**.

При решении практических задач зачастую точно найти закон распределения случайной величины довольно сложно. Однако, все происходящие процессы, связанные со случайными величинами, можно разделить на несколько типов, каждому из которых можно поставить в соответствие какой – либо закон распределения.

Выше были рассмотрены некоторые типы распределений дискретной случайной величины такие как биноминальное распределение и распределение Пуассона.

Рассмотрим теперь некоторые типы законов распределения для непрерывной случайной величины.

Лекция 9.

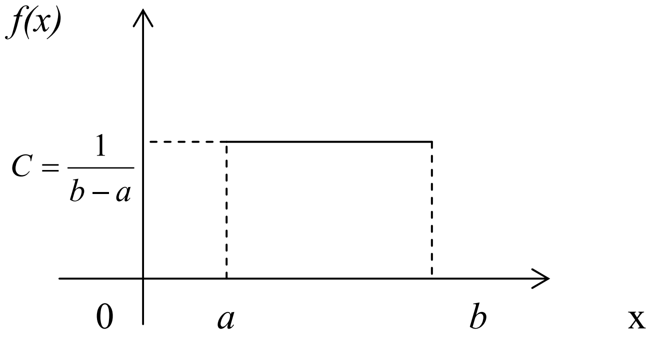
Равномерное распределение вероятностей. Показательное распределение. Функция надежности. Показательный закон надежности.

Равномерное распределение.

Определение. Непрерывная случайная величина имеет **равномерное** распределение на отрезке [*a, b*], если на этом отрезке плотность распределения случайной величины постоянна, а вне его равна нулю.

$$f(x) = \begin{cases} 0, & x < a \\ C, & a \le x \le b \\ 0, & x > b \end{cases}$$

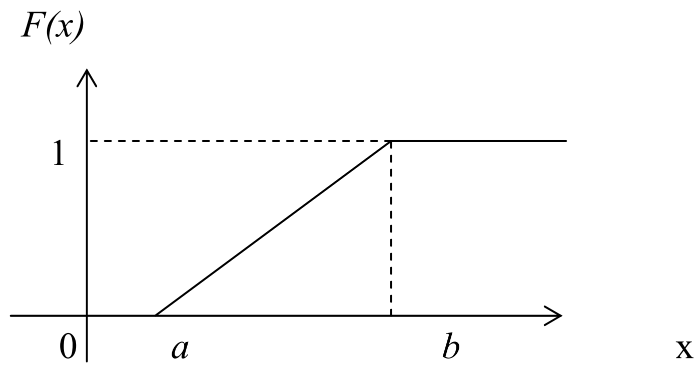
Постоянная величина С может быть определена из условия равенства единице площади, ограниченной кривой распределения.



Получаем $C = \frac{1}{b-a}$.

Найдем функцию распределения $F(x)$ на отрезке $[a,b]$.

$$F(x) = \begin{cases} 0, & при \quad x < a \\ \frac{x-a}{b-a}, & при \quad a \le x \le b \\ 1, & при \quad x > b \end{cases}$$



Для того, чтобы случайная величина подчинялась закону равномерного распределения необходимо, чтобы ее значения лежали внутри некоторого определенного интервала, и внутри этого интервала значения этой случайной величины были бы равновероятны.

Определим математическое ожидание и дисперсию случайной величины, подчиненной равномерному закону распределения.

$$m_x = \int_a^b xf(x)dx = \int_a^b \frac{x}{b-a}dx = \frac{x^2}{2(b-a)}\Big|_a^b = \frac{a+b}{2}.$$

$$m_{x^2} = \int_a^b x^2 f(x)dx = \int_a^b \frac{x^2}{b-a}dx = \frac{x^3}{3(b-a)}\Big|_a^b = \frac{b^3-a^3}{3(b-a)} = \frac{b^2+ab+a^2}{3}.$$

$$\sigma_x = \sqrt{D_x} = \frac{b-a}{2\sqrt{3}}.$$

Вероятность попадания случайной величины в заданный интервал:

$$P(\alpha < X < \beta) = \int_\alpha^\beta \frac{dx}{b-a} = \frac{\beta-\alpha}{b-a}.$$

Показательное распределение.

Определение. **Показательным (экспоненциальным)** называется распределение вероятностей непрерывной случайной величины X, которое описывается плотностью

$$f(x) = \begin{cases} 0, & \text{ïðè} \quad x < 0, \\ \lambda e^{-\lambda x}, & \text{ïðè} \quad x \geq 0, \end{cases}$$

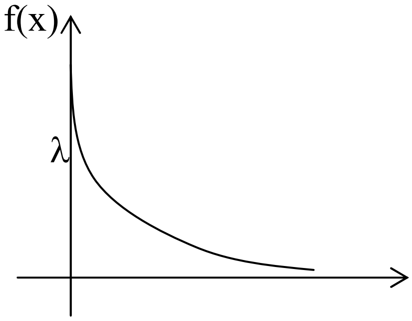
где λ – положительное число.

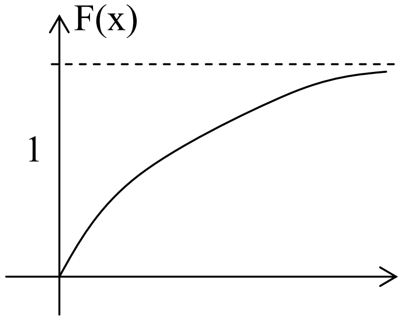
Найдем закон распределения.

$$F(x) = \int_{-\infty}^{x} f(x)dx = \int_{-\infty}^{0} 0dx + \lambda\int_0^x e^{-\lambda x}dx = 1 - e^{-\lambda x}.$$

$$F(x) = \begin{cases} 0, & \text{ïðè} \quad x < 0, \\ 1 - e^{-\lambda x} & \text{ïðè} \quad x \geq 0. \end{cases}$$

Графики функции распределения и плотности распределения:





0 x 0 x

Найдем математическое ожидание случайной величины, подчиненной показательному распределению.

$$m_x = \int_{-\infty}^{\infty} xf(x)dx = \int_{0}^{\infty} x\lambda e^{-\lambda x}dx = \begin{cases} u = x; & e^{-\lambda x}dx = dv; \\ du = dx; & -\dfrac{e^{-\lambda x}}{\lambda} = v; \end{cases} \Bigg\} = \lambda\left(-\frac{xe^{-\lambda x}}{\lambda}\Big|_0^{\infty} + \int_0^{\infty} \frac{e^{-\lambda x}}{\lambda}dx \right) =$$

$$= \int_0^{\infty} e^{-\lambda x}dx = -\frac{e^{-\lambda x}}{\lambda}\Big|_0^{\infty} = \frac{1}{\lambda}.$$

Результат получен с использованием того факта, что

$$xe^{-\lambda x}\Big|_0^{\infty} = \lim_{x\to\infty}\frac{x}{e^{\lambda x}} - 0 = \begin{Bmatrix} \text{По} \quad \text{правилу} \\ \text{Лопиталя} \end{Bmatrix} = \lim_{x\to\infty}\frac{1}{-\lambda e^{\lambda x}} = 0.$$

Для нахождения дисперсии найдем величину $M(X^2)$.

$$M(X^2) = \int_{-\infty}^{\infty} x^2 f(x)dx = \int_0^{\infty} \lambda x^2 e^{-\lambda x}dx$$

Дважды интегрируя по частям, аналогично рассмотренному случаю, получим: $M(X^2) = \dfrac{2}{\lambda^2}$;

Тогда $D(X) = M(X^2) - [M(X)]^2 = \dfrac{1}{\lambda^2}$.

Итого: $M(X) = \dfrac{1}{\lambda}$; $D(X) = \dfrac{1}{\lambda^2}$; $\sigma_x = \dfrac{1}{\lambda}$.

Видно, что в случае показательного распределения математическое ожидание и среднее квадратическое отклонение равны.

Также легко определить и вероятность попадания случайной величины, подчиненной показательному закону распределения, в заданный интервал. $P(a < x < b) = F(b) - F(a) = e^{-\lambda a} - e^{-\lambda b}$.

Показательное распределение широко используется в теории надежности.

Допустим, некоторое устройство начинает работать в момент времени $t_0=0$, а через какое-то время t происходит отказ устройства.

Обозначим T непрерывную случайную величину – длительность безотказной работы устройства.

Таким образом, функция распределения $F(t) = P(T<t)$ определяет вероятность отказа за время длительностью t.

Вероятность противоположного события (безотказная работа в течение времени t) равна $R(t) = P(T>t) = 1 – F(t)$.

Определение. **Функцией надежности** $R(t)$ называют функцию, определяющую вероятность безотказной работы устройства в течение времени t.

Часто на практике длительность безотказной работы подчиняется показательному закону распределению.

Вообще говоря, если рассматривать новое устройство, то вероятность отказа в начале его функционирования будет больше, затем количество отказов снизится и будет некоторое время иметь практически одно и то же значение. Затем (когда устройство выработает свой ресурс) количество отказов будет возрастать.

Другими словами, можно сказать, что функционирование устройства на протяжении всего существования (в смысле количества отказов) можно описать комбинацией двух показательных законов (в начале и конце функционирования) и равномерного закона распределения.

Функция надежности для какого- либо устройства при показательном законе распределения равна:

$$R(t) = 1 - F(t) = e^{-\lambda t}.$$

Данное соотношение называют **показательным законом надежности**.

Важным свойством, позволяющим значительно упростить решение задач теории надежности, является то, что вероятность безотказной работы устройства на интервале времени t не зависит от времени предшествующей работы до начала рассматриваемого интервала, а зависит только от длительности времени t.

Таким образом, безотказная работа устройства зависит только от интенсивности отказов λ и не зависит от безотказной работы устройства в прошлом.

Так как подобным свойством обладает только показательный закон распределения, то этот факт позволяет определить, является ли закон распределения случайной величины показательным или нет.

Лекция 10.

Нормальный закон распределения вероятностей. Нормальная кривая. Правило трех сигм.

Нормальный закон распределения.

Определение. **Нормальным** называется распределение вероятностей непрерывной случайной величины, которое описывается плотностью вероятности

$$f(x) = \frac{1}{\sigma_x \sqrt{2\pi}} e^{-\frac{(x-m_x)^2}{2\sigma_x^2}}.$$

Нормальный закон распределения также называется **законом Гаусса**.

Нормальный закон распределения занимает центральное место в теории вероятностей. Это обусловлено тем, что этот закон проявляется во всех случаях, когда случайная величина является результатом действия большого числа различных факторов. К нормальному закону приближаются все остальные законы распределения.

Можно легко показать, что параметры m_x и σ_x, входящие в плотность распределения являются соответственно математическим ожиданием и средним квадратическим отклонением случайной величины X.

Найдем функцию распределения *F(x)*.

$$F(x)=\frac{1}{\sigma_x\sqrt{2\pi}}\int_{-\infty}^{x} e^{-\frac{(x-m_x)^2}{2\sigma_x^2}}dx$$

График плотности нормального распределения называется **нормальной кривой** или **кривой Гаусса**.

Нормальная кривая обладает следующими свойствами:

1) Функция определена на всей числовой оси.

2) При всех x функция распределения принимает только положительные значения.

3) Ось OX является горизонтальной асимптотой графика плотности вероятности, т.к. при неограниченном возрастании по абсолютной величине аргумента x, значение функции стремится к нулю.

4) Найдем экстремум функции.

$$y'=-\frac{x-m}{\sigma^3\sqrt{2\pi}}e^{-\frac{(x-m)^2}{2\sigma^2}}=0; \qquad x=m;$$

Т.к. при $y' > 0$ при $x < m$ и $y' < 0$ при $x > m$, то в точке $x = m$ функция имеет максимум, равный $\frac{1}{\sigma\sqrt{2\pi}}$.

5) Функция является симметричной относительно прямой $x = a$, т.к. разность
$(x - a)$ входит в функцию плотности распределения в квадрате.

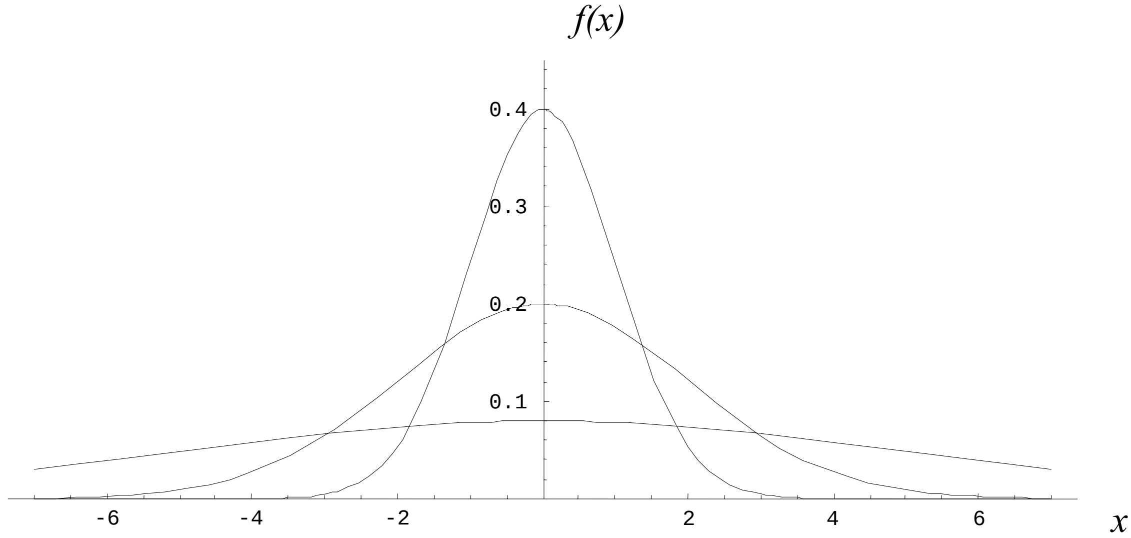
6) Для нахождения точек перегиба графика найдем вторую производную функции плотности.

$$y''=-\frac{1}{\sigma^3\sqrt{2\pi}}e^{-\frac{(x-m)^2}{2\sigma^2}}\left[1-\frac{(x-m)^2}{\sigma^2}\right]$$

При $x = m + \sigma$ и $x = m - \sigma$ вторая производная равна нулю, а при переходе через эти точки меняет знак, т.е. в этих точках функция имеет перегиб.

В этих точках значение функции равно $\frac{1}{\sigma e\sqrt{2\pi}}$.

Построим график функции плотности распределения.



Построены графики при m =0 и трех возможных значениях среднего квадратичного отклонения $\sigma = 1$, $\sigma = 2$ и $\sigma = 7$. Как видно, при увеличении значения среднего квадратичного отклонения график становится более пологим, а максимальное значение уменьшается.

Если $a > 0$, то график сместится в положительном направлении, если $a < 0$ – в отрицательном.

При $a = 0$ и $\sigma = 1$ кривая называется **нормированной**. Уравнение нормированной кривой:

$$\varphi(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}.$$

Найдем вероятность попадания случайной величины, распределенной по нормальному закону, в заданный интервал.

$$P(a < X < b) = \int_a^b f(x)dx = \frac{1}{\sigma\sqrt{2\pi}} \int_a^b e^{-\frac{(x-m)^2}{2\sigma^2}} dx$$

Обозначим $\frac{x-m}{\sigma\sqrt{2}} = t;\quad \frac{a-m}{\sigma\sqrt{2}} = \alpha;\quad \frac{b-m}{\sigma\sqrt{2}} = \beta;$

Тогда $P(a < X < b) = \frac{1}{\sigma\sqrt{2\pi}}\int\limits_{\alpha}^{\beta} e^{-t^2}\sigma\sqrt{2}dt = \frac{1}{\sqrt{\pi}}\int\limits_{\alpha}^{\beta} e^{-t^2}dt = \frac{1}{2}\left[\Phi(\beta) - \Phi(\alpha)\right]$

Т.к. интеграл $\int e^{-t^2}dt$ не выражается через элементарные функции, то вводится в рассмотрение функция

$$\Phi(x) = \frac{2}{\sqrt{\pi}}\int\limits_{0}^{x} e^{-t^2}dt,$$

которая называется **функцией Лапласа** или **интегралом вероятностей**.

При рассмотрении нормального закона распределения выделяется важный частный случай, известный как **правило трех сигм**. Запишем вероятность того, что отклонение нормально распределенной случайной величины от математического ожидания меньше заданной величины

$\Delta\, P(|X - m| < \Delta) = \overline{\Phi}\left[\frac{m + \Delta - m}{\sigma}\right] - \overline{\Phi}\left[\frac{m - \Delta - m}{\sigma}\right] = \overline{\Phi}\left[\frac{\Delta}{\sigma}\right] - \overline{\Phi}\left[-\frac{\Delta}{\sigma}\right] = 2\overline{\Phi}\left[\frac{\Delta}{\sigma}\right]$

Если принять $\Delta = 3\sigma$, то получаем с использованием таблиц значений функции Лапласа:

$$P(|X - m| < 3\sigma) = 2\overline{\Phi}(3) = 2 \cdot 0{,}49865 = 0{,}9973$$

Т.е. вероятность того, что случайная величина отклонится от своего математического ожидание на величину, большую чем утроенное среднее квадратичное отклонение, практически равна нулю.

Это правило называется **правилом трех сигм**.

Не практике считается, что если для какой – либо случайной величины выполняется правило трех сигм, то эта случайная величина имеет нормальное распределение.

<u>Центральная предельная теорема Ляпунова.</u>

<u>Теорема.</u> *Если случайная величина X представляет собой сумму очень большого числа взаимно независимых случайных величин, влияние каждой из* *которых на всю сумму ничтожно мало, то X имеет распределение, близкое к нормальному.*

На практике для большинства случайных величин выполняются условия теоремы Ляпунова.

Лекция 11. Система случайных величин. Плотность распределения системы двух случайных величин. Условные законы распределения. Зависимые и независимые случайные величины. Линейная корреляция.

Система случайных величин.

Рассмотренные выше случайные величины были одномерными, т.е. определялись одним числом, однако, существуют также случайные величины, которые определяются двумя, тремя и т.д. числами. Такие случайные величины называются двумерными, трехмерными и т.д.

В зависимости от типа, входящих в систему случайных величин, системы могут быть дискретными, непрерывными или смешанными, если в систему входят различные типы случайных величин.

Более подробно рассмотрим системы двух случайных величин.

Определение. **Законом распределения** системы случайных величин называется соотношение, устанавливающее связь между областями возможных значений системы случайных величин и вероятностями появления системы в этих областях.

Определение. **Функцией распределения** системы двух случайных величин называется функция двух аргументов *F(x, y)*, равная вероятности совместного выполнения двух неравенств $X<x,\ Y<y$.

$$F(x, y) = P(X < x, Y < y)$$

Отметим следующие свойства функции распределения системы двух случайных величин:

1) Если один из аргументов стремится к плюс бесконечности, то функция распределения системы стремится к функции распределения одной случайной величины, соответствующей другому аргументу.

$$\lim_{y\to\infty} F(x,y) = F_1(x);$$

$$\lim_{x\to\infty} F(x,y) = F_2(y);$$

2) Если оба аргумента стремятся к бесконечности, то функция распределения системы стремится к единице.

$$F(\infty,\infty) = 1;$$

3) При стремлении одного или обоих аргументов к минус бесконечности функция распределения стремится к нулю.

$$F(-\infty, y) = F(x,-\infty) = F(-\infty,-\infty) = 0;$$

4) Функция распределения является неубывающей функцией по каждому аргументу.

5) Вероятность попадания случайной точки (X, Y) в произвольный прямоугольник со сторонами, параллельными координатным осям, вычисляется по формуле:

$$P(a \le X \le b, c \le Y \le d) = F(b,d) - F(a,d) - F(b,c) + F(a,c)$$

Плотность распределения системы двух случайных величин.

Определение. Плотностью совместного распределения вероятностей двумерной случайной величины (X, Y) называется вторая смешанная частная производная от функции распределения.

$$f(x,y) = \frac{\partial^2 F(x,y)}{\partial x \partial y}$$

Если известна плотность распределения, то функция распределения может быть легко найдена по формуле:

$$F(x,y) = \int_{-\infty}^{y}\int_{-\infty}^{x} f(x,y)dxdy$$

Двумерная плотность распределения неотрицательна и двойной интеграл с бесконечными пределами от двумерной плотности равен единице.

$$\int\limits_{-\infty}^{\infty}\int\limits_{-\infty}^{\infty} f(x,y)dxdy = 1$$

По известной плотности совместного распределения можно найти плотности распределения каждой из составляющих двумерной случайной величины.

$$f_1(x) = \int\limits_{-\infty}^{\infty} f(x,y)dy; \qquad f_2(y) = \int\limits_{-\infty}^{\infty} f(x,y)dx$$

$$F_1(x) = F(x,\infty) = \int\limits_{-\infty}^{x}\int\limits_{-\infty}^{\infty} f(x,y)dxdy\,; \quad F_2(y) = F(\infty,y) = \int\limits_{-\infty}^{y}\int\limits_{-\infty}^{\infty} f(x,y)dxdy\,.$$

<u>Условные законы распределения</u>.

Как было показано выше, зная совместный закон распределения можно легко найти законы распределения каждой случайной величины, входящей в систему.

Однако, на практике чаще стоит обратная задача – по известным законам распределения случайных величин найти их совместный закон распределения.

В общем случае эта задача является неразрешимой, т.к. закон распределения случайной величины ничего не говорит о связи этой величины с другими случайными величинами.

Кроме того, если случайные величины зависимы между собой, то закон распределения не может быть выражен через законы распределения составляющих, т.к. должен устанавливать связь между составляющими.

Все это приводит к необходимости рассмотрения условных законов распределения.

<u>**Определение.**</u> Распределение одной случайной величины, входящей в систему, найденное при условии, что другая случайная величина приняла определенное значение, называется **условным законом распределения**.

Условный закон распределения можно задавать как функцией распределения так и плотностью распределения.

Условная плотность распределения вычисляется по формулам:

$$f(x/y)=\frac{f(x,y)}{f_2(y)}=\frac{f(x,y)}{\int\limits_{-\infty}^{\infty}f(x,y)dx}$$

$$f(y/x)=\frac{f(x,y)}{f_1(x)}=\frac{f(x,y)}{\int\limits_{-\infty}^{\infty}f(x,y)dy}$$

Условная плотность распределения обладает всеми свойствами плотности распределения одной случайной величины.

Условное математическое ожидание.

Определение. Условным математическим ожиданием дискретной случайной величины Y при X = x (x – определенное возможное значение X) называется произведение всех возможных значений Y на их условные вероятности.

$$M(Y/X=x)=\sum_{i=1}^{m}y_i p(y_i/x)$$

Для непрерывных случайных величин:

$$M(Y/X=x)=\int\limits_{-\infty}^{\infty}yf(y/x)dy\,,$$

где *f(y/x)* – условная плотность случайной величины Y при X=x.

Условное математическое ожидание *M(Y/x)=f(x)* является функцией от *x* и называется **функцией регрессии X на Y.**

Аналогично определяются условная дисперсия и условные моменты системы случайных величин.

Зависимые и независимые случайные величины.

Случайные величины называются независимыми, если закон распределения одной из них не зависит от того, какое значение принимает другая случайная величина.

Понятие зависимости случайных величин является очень важным в теории вероятностей.

Условные распределения независимых случайных величин равны их безусловным распределениям.

Определим необходимые и достаточные условия независимости случайных величин.

Теорема. *Для того, чтобы случайные величины X и Y были независимы, необходимо и достаточно, чтобы функция распределения системы (X, Y) была равна произведению функций распределения составляющих.*

$$F(x, y) = F_1(x)F_2(y)$$

Аналогичную теорему можно сформулировать и для плотности распределения:

Теорема. *Для того, чтобы случайные величины X и Y были независимы, необходимо и достаточно, чтобы плотность совместного распределения системы (X, Y) была равна произведению плотностей распределения составляющих.*

$$f(x, y) = f_1(x)f_2(y)$$

Определение. Корреляционным моментом μ_{xy} случайных величин X и Y называется математическое ожидание произведения отклонений этих величин.

$$\mu_{xy} = M\{[X - M(X)][Y - M(Y)]\}$$

Практически используются формулы:

Для дискретных случайных величин:

$$\mu_{xy} = \sum_{i=1}^{n}\sum_{j=1}^{m}[x_i - M(X)][y_j - M(Y)]p(x_i, y_j)$$

Для непрерывных случайных величин:

$$\mu_{xy} = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}[x - M(X)][y - M(Y)]f(x,y)dxdy$$

Корреляционный момент служит для того, чтобы охарактеризовать связь между случайными величинами. Если случайные величины независимы, то их корреляционный момент равен нулю.

Корреляционный момент имеет размерность, равную произведению размерностей случайных величин X и Y. Этот факт является недостатком этой числовой характеристики, т.к. при различных единицах измерения получаются различные корреляционные моменты, что затрудняет сравнение корреляционных моментов различных случайных величин.

Для того, чтобы устранить этот недостаток применятся другая характеристика – коэффициент корреляции.

Определение. **Коэффициентом корреляции** $\boldsymbol{r_{xy}}$ случайных величин X и Y называется отношение корреляционного момента к произведению средних квадратических отклонений этих величин.

$$r_{xy} = \frac{\mu_{xy}}{\sigma_x \sigma_y}$$

Коэффициент корреляции является безразмерной величиной. Коэффициент корреляции независимых случайных величин равен нулю.

Свойство: Абсолютная величина корреляционного момента двух случайных величин X и Y не превышает среднего геометрического их дисперсий.

$$|\mu_{xy}| \le \sqrt{D_x D_y}$$

Свойство: Абсолютная величина коэффициента корреляции не превышает единицы.

$$|r_{xy}| \le 1$$

Случайные величины называются **коррелированными**, если их корреляционный момент отличен от нуля, и **некоррелированными**, если их корреляционный момент равен нулю.

Если случайные величины независимы, то они и некоррелированы, но из некоррелированности нельзя сделать вывод о их независимости.

Если две величины зависимы, то они могут быть как коррелированными, так и некоррелированными.

Часто по заданной плотности распределения системы случайных величин можно определить зависимость или независимость этих величин.

Наряду с коэффициентом корреляции степень зависимости случайных величин можно охарактеризовать и другой величиной, которая называется **коэффициентом ковариации**. Коэффициент ковариации определяется формулой:

$$k(X,Y) = M(XY) - M(X) \cdot M(Y)$$

<u>Линейная регрессия.</u>

Рассмотрим двумерную случайную величину (X, Y), где X и Y – зависимые случайные величины. Представим приближенно одну случайную величину как функцию другой. Точное соответствие невозможно. Будем считать, что эта функция линейная.

$$Y \cong g(X) = \alpha X + \beta$$

Для определения этой функции остается только найти постоянные величины α и β.

<u>Определение.</u> Функция *g(X)* называется **наилучшим приближением** случайной величины Y в смысле метода наименьших квадратов, если математическое ожидание

$M[Y - g(X)]^2$ принимает наименьшее возможное значение. Также функция *g(x)* называется **среднеквадратической регрессией** Y на X.

<u>Теорема.</u> *Линейная средняя квадратическая регрессия Y на X вычисляется по формуле:* $g(X) = m_y + r\dfrac{\sigma_y}{\sigma_x}(X - m_x)$

в этой формуле m_x=*M(X)*, m_y=*M(Y)*, $\sigma_x = \sqrt{D(X)}$, $\sigma_y = \sqrt{D(Y)}$, $r = \mu_{xy}/(\sigma_x\sigma_y)$ – коэффициент корреляции величин *X* и *Y*.

Величина $\beta = r\frac{\sigma_y}{\sigma_x}$ называется **коэффициентом регрессии** *Y* на *X*.

Прямая, уравнение которой

$$y - m_y = r\frac{\sigma_y}{\sigma_x}(x - m_x),$$

называется **прямой сренеквадратической регрессии** *Y* на *X*.

Величина $\sigma_y^2(1-r^2)$ называется **остаточной дисперсией** случайной величины *Y* относительно случайной величины *X*. Эта величина характеризует величину ошибки, образующейся при замене случайной величины *Y* линейной функцией $g(X)=\alpha X + \beta$.

Видно, что если $r=\pm 1$, то остаточная дисперсия равна нулю, и, следовательно, ошибка равна нулю и случайная величина *Y* точно представляется линейной функцией от случайной величины *X*.

Прямая среднеквадратичной регрессии *X* на *Y* определяется аналогично по формуле:

$$x - m_x = r\frac{\sigma_x}{\sigma_y}(y - m_y)$$

Прямые среднеквадратичной регрессии пересекаются в точке (m_x, m_y), которую называют **центром совместного распределения** случайных величин *X* и *Y*.

Линейная корреляция.

Если две случайные величины *X* и *Y* имеют в отношении друг друга линейные функции регрессии, то говорят, что величины *X* и *Y* связаны **линейной корреляционной зависимостью**.

Теорема. *Если двумерная случайная величина (X, Y) распределена нормально, то X и Y связаны линейной корреляционной зависимостью.*

Лекция 12.

Закон больших чисел. Неравенство Чебышева. Теоремы Чебышева и Бернулли.

Закон больших чисел.

На практике сложно сказать какое конкретное значение примет случайная величина, однако, при воздействии большого числа различных факторов поведение большого числа случайных величин практически утрачивает случайный характер и становится закономерным.

Этот факт очень важен на практике, т.к. позволяет предвидеть результат опыта при воздействии большого числа случайных факторов.

Однако, это возможно только при выполнении некоторых условий, которые определяются законом больших чисел. К законам больших чисел относятся теоремы Чебышева (наиболее общий случай) и теорема Бернулли (простейший случай), которые будут рассмотрены далее.

Рассмотрим дискретную случайную величину *X* (хотя все сказанное ниже будет справедливо и для непрерывных случайных величин), заданную таблицей распределения:

X	x_1	x_2	…	x_n
p	p_1	p_2	…	p_n

Требуется определить вероятность того, что отклонение значения случайной величины от ее математического ожидания будет не больше, чем заданное число ε.

Теорема. (Неравенство Чебышева) *Вероятность того, что отклонение случайной величины X от ее математического ожидания по абсолютной величине меньше положительного числа ε, не меньше чем* $1 - D(X)/\varepsilon^2$.

$$P(|X - M(X)| < \varepsilon) \geq 1 - D(X)/\varepsilon^2$$

Доказательство этой теоремы приводить не будем, оно имеется в литературе.

Теорема Чебышева.

Теорема. *Если X_1, X_2, …, X_n- попарно независимые случайные величины, причем дисперсии их равномерно ограничены (не превышаю постоянного числа С), то, как бы мало не было положительное число ε, вероятность неравенства*

$$\left[\frac{X_1+X_2+...+X_n}{n}-\frac{M(X_1)+M(X_2)+...+M(X_n)}{n}\right]<\varepsilon$$

будет сколь угодно близка к единице, если число случайных величин достаточно велико.

Т.е. можно записать:

$$\lim_{n\to\infty}P\left(\left|\frac{X_1+X_2+...+X_n}{n}-\frac{M(X_1)+M(X_2)+...+M(X_n)}{n}\right|<\varepsilon\right)=1$$

Часто бывает, что случайные величины имеют одно и то же математическое ожидание. В этом случае теорема Чебышева несколько упрощается:

$$\lim_{n\to\infty}P\left(\left|\frac{X_1+X_2+...+X_n}{n}-m\right|<\varepsilon\right)=1$$

Дробь, входящая в записанное выше выражение есть не что иное как среднее арифметическое возможных значений случайной величины.

Теорема утверждает, что хотя каждое отдельное значение случайной величины может достаточно сильно отличаться от своего математического ожидания, но среднее арифметическое этих значений будет неограниченно приближаться к среднему арифметическому математических ожиданий.

Отклоняясь от математического ожидания как в положительную так и в отрицательную сторону, от своего математического ожидания, в среднем арифметическом отклонения взаимно сокращаются.

Таким образом, величина среднего арифметического значений случайной величины уже теряет характер случайности.

Теорема Бернулли.

Пусть производится n независимых испытаний, в каждом из которых вероятность появления события А равна p.

Возможно определить примерно относительную частоту появления события А.

Теорема. *Если в каждом из n независимых испытаний вероятность p появления события A постоянно, то сколь угодно близка к единице вероятность того, что отклонение относительной частоты от вероятности p по абсолютной величине будет сколь угодно малым, если число испытаний p достаточно велико.*

$$\lim_{n\to\infty} P(|m/n - p| < \varepsilon) = 1$$

Здесь m – число появлений события А. Из всего сказанного выше не следует, что с увеличением число испытаний относительная частота неуклонно стремится к вероятности p, т.е. $\lim\limits_{n\to\infty}\frac{m}{n} = p$. В теореме имеется в виду только вероятность приближения относительной частоты к вероятности появления события А в каждом испытании.

В случае, если вероятности появления события А в каждом опыте различны, то справедлива следующая теорема, известная как **теорема Пуассона.**

Теорема. *Если производится n независимых опытов и вероятность появления события A в каждом опыте равна* p_i*, то при увеличении n частота события A сходится по вероятности к среднему арифметическому вероятностей* p_i.

Предельные теоремы.

Как уже говорилось, при достаточно большом количестве испытаний, поставленных в одинаковых условиях, характеристики случайных событий и случайных величин становятся почти неслучайными. Это позволяет использовать результаты наблюдений случайных событий для предсказания исхода того или иного опыта.

Здесь m – число появлений события А. Из всего сказанного выше не следует, что с увеличением число испытаний относительная частота неуклонно стремится к вероятности p, т.е. $\lim\limits_{n\to\infty}\frac{m}{n}=p$. В теореме имеется в виду только вероятность приближения относительной частоты к вероятности появления события А в каждом испытании.

В случае, если вероятности появления события А в каждом опыте различны, то справедлива следующая теорема, известная как **теорема Пуассона.**

Теорема. *Если производится n независимых опытов и вероятность появления события A в каждом опыте равна p_i, то при увеличении n частота события A сходится по вероятности к среднему арифметическому вероятностей p_i.*

Предельные теоремы.

Как уже говорилось, при достаточно большом количестве испытаний, поставленных в одинаковых условиях, характеристики случайных событий и случайных величин становятся почти неслучайными. Это позволяет использовать результаты наблюдений случайных событий для предсказания исхода того или иного опыта.

Предельные теоремы теории вероятностей устанавливают соответствие между теоретическими и экспериментальными характеристиками случайных величин при большом количестве испытаний.

В рассмотренном выше законе больших чисел ничего не говорилось о законе распределения случайных величин.

Поставим задачу нахождения предельного закона распределения суммы $Y_n = \sum_{i=1}^{n} X_i$ когда число слагаемых n неограниченно возрастает. Эту задачу решает Центральная предельная теорема Ляпунова, которая была сформулирована выше.

В зависимости от условий распределения случайных величин X_i, образующих сумму, возможны различные формулировки центральной предельной теоремы.

Допустим, что случайные величины X_i взаимно независимы и одинаково распределены.

Теорема. *Если случайные величины X_i взаимно независимы и имеют один и тот же закон распределения с математическим ожиданием m и дисперсией σ^2, причем существует третий абсолютный момент ν_3, то при неограниченном увеличении числа испытаний n закон распределения суммы* $Y_n = \sum_{i=1}^{n} X_i$ *неограниченно приближается к нормальному.*

Лекция 13 .

Теория массового обслуживания. Случайные процессы. Поток событий. Цепи Маркова.

Система массового обслуживания состоит из некоторого числа обслуживающих единиц или **каналов**, работа которых состоит в выполнении поступающих по этим каналам **заявок**.

Примеры систем массового обслуживания весьма распространены на практике. Это различные телефонные станции, ремонтные мастерские и прочее. Вид и количество поступающих на эти системы заявок различны и, вообще говоря, случайны.

Теория массового обслуживания описывает закономерности функционирования таких систем.

<u>Определение.</u> Процесс функционирования системы массового обслуживания называется **случайным процессом.**

Чтобы оптимизировать процесс функционирования системы массового обслуживания его надо изучить и описать математически.

Теория массового обслуживания является очень быстро развивающимся разделом теории вероятностей, т.к. ее применение на практике чрезвычайно широко.

Случайный процесс, протекающий в системе массового обслуживания, состоит в том, что система в случайные моменты времени переходит из одного состояния в другое. Меняется число заявок, число занятых каналов, число заявок в очереди и проч.

<u>Определение.</u> Если переход системы из одного состояния в другое происходит скачком, а количество состояний системы (конечное или бесконечное) можно пронумеровать, то такая система называется **системой дискретного типа.**

Если количество возможных состояний счетно, то сумма вероятностей нахождения системы в одном из состояний равна 1.

$$\sum_k p_k(t) = 1$$

Совокупность вероятностей $p_k(t)$ для каждого момента времени характеризует данное **сечение** случайного процесса.

Случайные процессы со счетным множеством состояний бывают двух типов: **с дискретным** или **непрерывным временем**.

Если переходы системы из одного состояния в другое могут происходить только в строго определенные моменты времени, то случайный процесс будет процессом с дискретным временем, а если переход возможен в любой момент времени, то процесс будет процессом с непрерывным временем.

Поскольку в реальности заявки на систему массового обслуживания могут поступать в любой момент времени, то большинство реальных систем массового обслуживания будут системами с процессом с непрерывным временем.

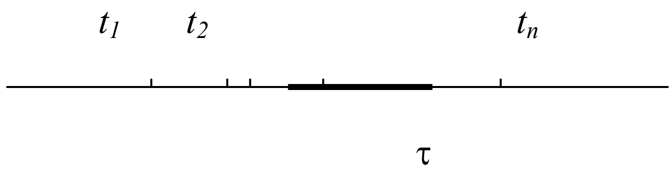
Для того, чтобы описать случайный процесс в системе с непрерывным временем необходимо, прежде всего, проанализировать причины, вызывающие изменение состояния системы. Эти причины определяются потоком заявок, поступающих на систему.

Поток событий.

Определение. **Потоком событий** называется последовательность событий, происходящих один за другим в какие-то моменты времени.

Характер событий, образующих поток может быть различным, а если события отличаются друг от друга только моментом времени, в который они происходят, то такой поток событий называется **однородным**.

Однородный поток можно изобразить последовательностью точек на оси, соответствующей времени:



Определение. Поток событий называется **регулярным**, если события следует одно за другим через строго определенные промежутки времени.

Определение. Поток событий называется **стационарным**, если вероятность попадания того ли иного числа событий на участок времени τ зависит только от длины участка и не зависит от того, где именно на оси расположен этот участок.

Стационарность потока событий означает, что плотность потока постоянна, отсутствуют промежутки времени, в течение которых событий больше чем обычно. Классический пример – “час пик” на транспорте.

Определение. Поток событий называется **потоком без последействий**, если для любых неперекрещивающихся участков времени число событий, попадающих на один из них, не зависит от числа событий, попадающих на другие.

Отсутствие последействий означает, что заявки в систему поступают независимо друг от друга. Поток выходных событий систем массового обслуживания обычно имеет последействие, даже если входной поток его не имеет. Пример – вход пассажиров на станцию метро – поток без последействия, т.к. причины прихода отдельного пассажира не связаны с причинами прихода всех остальных, а выход пассажиров со станции – поток с последействием, т.к. он обусловлен прибытием поезда.

Последействие, свойственное выходному потоку следует учитывать, если этот поток в свою очередь является входным для какой- либо другой системы.

Определение. Поток событий называется **ординарным**, если вероятность попадания на элементарный участок Δt двух или более событий достаточно мало по сравнению с вероятностью попадания одного события.

Условие ординарности означает, что заявки на систему приходят по одному, а не парами, тройками и т.д. Однако, если заявки поступают **только** парами, **только** тройками и т.д., то такой поток легко свести к ординарному.

Определение. Если поток событий стационарен, ординарен и без последействий, то такой поток называется **простейшим (пуассоновским)** потоком.

Это название связано с тем, что в этом случае число событий, попадающих на любой фиксированный интервал времени, распределено по распределению Пуассона.

В соответствии с этим законом распределения математическое ожидание числа точек, попавших попадающих на участок времени τ, имеет вид:

$$a = \lambda\tau$$

λ - плотность потока – среднее число событий в единицу времени.

Вероятность того, что за время τ произойдет ровно m событий, равна

$$P_m(\tau) = \frac{(\lambda\tau)^m}{m!} e^{-\lambda\tau}$$

Вероятность того, что в течение данного времени не произойдет ни одного события, равна:

$$P_0(\tau) = e^{-\lambda\tau}$$

Пусть Т – промежуток времени между двумя произвольными соседними событиями в простейшем потоке. Найдем функцию распределения

$$F(t) = P(T < t)$$

В соответствии с законом распределения Пуассона, получаем:

$$F(t) = 1 - e^{-\lambda t}; \qquad f(t) = \lambda e^{-\lambda t};$$

Математическое ожидание, дисперсия и среднее квадратическое отклонение этой величины соответственно равны:

$$m_t = \frac{1}{\lambda}; \quad D_t = \frac{1}{\lambda^2}; \quad \sigma_t = \frac{1}{\lambda};$$

Таким образом, для величины Т получили показательный закон распределения.

Пример. В бюро обслуживания в среднем поступает 12 заявок в час. Считая поток заказов простейшим, определить вероятность того, что: а) за 1 минуту не поступит ни одного заказа, б) за 10 минут поступит не более трех заказов.

Сначала найдем плотность (интенсивность) потока, выразив ее в количестве заявок в минуту. Очевидно, эта величина равна $\lambda = \frac{12}{60} = 0{,}2$.

Далее находим вероятность того, что за время $\tau = 1$ мин не поступит ни одной заявки по формуле:

$$P_0(\tau) = e^{-\lambda\tau} = e^{-0,2} \approx 0{,}819$$

Вероятность того, что за 10 минут поступит не более трех заказов будет складываться из вероятностей того, что не поступит ни одного заказа, поступит один, два или ровно три заказа.

$$P(m \le 3) = \sum_{m=0}^{3} \frac{(\lambda\tau)^m}{m!} e^{-\lambda\tau} = e^{-2} + 2e^{-2} + \frac{4}{2}e^{-2} + \frac{8}{6}e^{-2} = \frac{19}{3}e^{-2} = 0{,}8571$$

Пример. В ресторан прибывает в среднем 20 посетителей в час. Считая поток посетителей простейшим, и зная, что ресторан открывается в 11.00, определите:

а) вероятность того, что в 11.12 в ресторан придет 20 посетителей при условии, что в 11.07 их было 18

б) вероятность того, что между 11.28 и 11.30 в ресторане окажется новый посетитель, если известно, что предшествующий посетитель прибыл в 11.25.

Для ответ на первый вопрос фактически надо найти вероятность того, что в промежуток от 11.07 до 11.12 ($\tau = 5$ минут) придет ровно 2 посетителя. При этом мы знаем интенсивность потока посетителей - $\lambda = 20/60 = 1/3$ посетителей в минуту. Конечно, данная величина носит условный характер, т.к. посетители не могут приходить по частям.

Искомая вероятность равна:

$$P_2(5) = \frac{\left(\frac{1}{3}\cdot 5\right)^2}{2!} e^{-\frac{1}{3}\cdot 5} \approx 0{,}2623.$$

Цепи Маркова.

Определение. Процесс, протекающий в физической системе, называется **марковским**, если в любой момент времени вероятность любого состояния системы в будущем зависит только от состояния системы в текущий момент и не зависит от того, каким образом система пришла в это состояние.

Определение. **Цепью Маркова** называется последовательность испытаний, в каждом из которых появляется только одно из k несовместных событий A_i из полной группы. При этом условная вероятность $p_{ij}(s)$ того, что в s –ом испытании наступит событие A_j при условии, что в $(s-1)$ – ом испытании наступило событие A_i, не зависит от результатов предшествующих испытаний.

Независимые испытания являются частным случаем цепи Маркова. События называются **состояниями системы**, а испытания – **изменениями состояний системы**.

По характеру изменений состояний цепи Маркова можно разделить на две группы.

Определение. **Цепью Маркова с дискретным временем** называется цепь, изменение состояний которой происходит в определенные фиксированные моменты времени. **Цепью Маркова с непрерывным временем** называется цепь, изменение состояний которой возможно в любые случайные моменты времени.

Определение. **Однородной** называется цепь Маркова, если условная вероятность p_{ij} перехода системы из состояния i в состояние j не зависит от номера испытания. Вероятность p_{ij} называется **переходной вероятностью**.

Допустим, число состояний конечно и равно k.
Тогда матрица, составленная из условных вероятностей перехода будет иметь вид:

$$P_1 = \begin{pmatrix} p_{11} & p_{12} & \cdots & p_{1k} \\ p_{21} & p_{22} & \cdots & p_{2k} \\ \cdots & \cdots & \cdots & \cdots \\ p_{k1} & p_{k2} & \cdots & p_{kk} \end{pmatrix}$$

Эта матрица называется **матрицей перехода системы**.

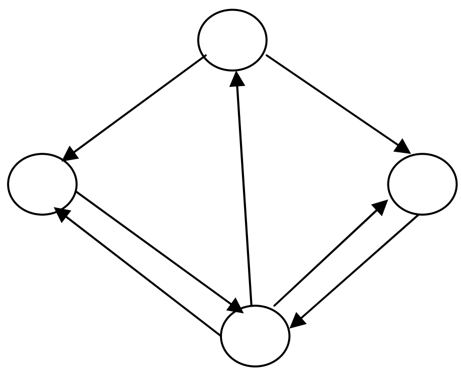
Т.к. в каждой строке содержаться вероятности событий, которые образуют полную группу, то, очевидно, что сумма элементов каждой строки матрицы равна единице.

На основе матрицы перехода системы можно построить так называемый **граф состояний системы**, его еще называют **размеченный граф состояний**. Это удобно для наглядного представления цепи. Порядок построения граф рассмотрим на примере.

Пример. По заданной матрице перехода построить граф состояний.

$$P_1 = \begin{pmatrix} 0{,}1 & 0{,}2 & 0 & 0{,}7 \\ 0 & 0{,}4 & 0{,}6 & 0 \\ 0{,}4 & 0{,}1 & 0 & 0{,}5 \\ 0 & 0 & 0{,}5 & 0{,}5 \end{pmatrix}$$

Т.к. матрица четвертого порядка, то, соответственно, система имеет 4 возможных состояния.



На графе не отмечаются вероятности перехода системы из одного состояния в то же самое. При рассмотрении конкретных систем удобно сначала построить граф состояний, затем определить вероятность переходов системы из одного состояния в то же самое (исходя из требования равенства единице суммы элементов строк матрицы), а потом составить матрицу переходов системы.

Пусть $P_{ij}(n)$ – вероятность того, что в результате n испытаний система перейдет из состояния i в состояние j, r – некоторое промежуточное состояние между состояниями i и j. Вероятности перехода из одного состояния в другое $p_{ij}(1) = p_{ij}$.

Тогда вероятность $P_{ij}(n)$ может быть найдена по формуле, называемой **равенством Маркова**:

$$P_{ij}(n)=\sum_{r=1}^{k}P_{ir}(m)P_{rj}(n-m)$$

Здесь m – число шагов (испытаний), за которое система перешла из состояния i в состояние r.

В принципе, равенство Маркова есть ни что иное, как несколько видоизмененная формула полной вероятности.

Зная переходные вероятности (т.е. зная матрицу перехода P_1), можно найти вероятности перехода из состояния в состояние за два шага $P_{ij}(2)$, т.е. матрицу P_2, зная ее – найти матрицу P_3, и т.д.

Непосредственное применений полученной выше формулы не очень удобно, поэтому, можно воспользоваться приемами матричного исчисления (ведь эта формула по сути – не что иное как формула перемножения двух матриц).

Тогда в общем виде можно записать:

$$P_n = P_1^n$$

<u>Пример.</u> Задана матрица переходов P_1. Найти матрицу P_3.

$$P_1=\begin{pmatrix}0{,}1 & 0{,}9\\ 0{,}3 & 0{,}7\end{pmatrix}$$

$$P_2=\begin{pmatrix}0{,}1 & 0{,}9\\ 0{,}3 & 0{,}7\end{pmatrix}\begin{pmatrix}0{,}1 & 0{,}9\\ 0{,}3 & 0{,}7\end{pmatrix}=\begin{pmatrix}0{,}28 & 0{,}72\\ 0{,}24 & 0{,}76\end{pmatrix}$$

$$P_3=\begin{pmatrix}0{,}1 & 0{,}9\\ 0{,}3 & 0{,}7\end{pmatrix}\begin{pmatrix}0{,}28 & 0{,}72\\ 0{,}24 & 0{,}76\end{pmatrix}=\begin{pmatrix}0{,}244 & 0{,}756\\ 0{,}252 & 0{,}748\end{pmatrix}$$

<u>Определение.</u> Матрицы, суммы элементов всех строк которых равны единице, называются **стохастическими**. Если при некотором n все элементы матрицы P^n не равны нулю, то такая матрица переходов называется **регулярной**.

Другими словами, регулярные матрицы переходов задают цепь Маркова, в которой каждое состояние может быть достигнуто через n шагов из любого состояния. Такие цепи Маркова также называются **регулярными**.

<u>Теорема.</u> (теорема о предельных вероятностях) *Пусть дана регулярная цепь Маркова с n состояниями и P – ее матрица вероятностей перехода. Тогда существует предел* $\lim_{n\to\infty} P^n = P^{(\infty)}$ *и матрица* $P^{(\infty)}$ *имеет вид:*

$$P^{(\infty)} = \begin{pmatrix} u_1 & u_2 & \dots & u_n \\ u_1 & u_2 & \dots & u_n \\ \dots & \dots & \dots & \dots \\ u_1 & u_2 & \dots & u_n \end{pmatrix}$$

Т.е. матрица состоит из одинаковых строк.

Теперь о величинах u_i. Числа $u_1, u_2, \dots, u_n$ называются **предельными вероятностями.** Эти вероятности не зависят от исходного состояния системы и являются компонентами собственного вектора матрицы P^T (транспонированной к матрице P).

Этот вектор полностью определяется из условий:

$$P^T \cdot \vec{u} = \vec{u}; \qquad \sum u_i = 1;$$

<u>Пример.</u> Найдем предельные вероятности для рассмотренного <u>выше</u> примера.

$$P = \begin{pmatrix} 0{,}1 & 0{,}9 \\ 0{,}3 & 0{,}7 \end{pmatrix}; \qquad P^T = \begin{pmatrix} 0{,}1 & 0{,}3 \\ 0{,}9 & 0{,}7 \end{pmatrix}$$

$$\begin{pmatrix} 0{,}1 & 0{,}3 \\ 0{,}9 & 0{,}7 \end{pmatrix} \begin{pmatrix} u_1 \\ u_2 \end{pmatrix} = \begin{pmatrix} u_1 \\ u_2 \end{pmatrix}$$

$$u_1 = 0{,}1u_1 + 0{,}3u_2$$
$$u_2 = 0{,}9u_1 + 0{,}7u_2$$

$$0{,}9u_1 = 0{,}3u_2; \qquad u_2 = 3u_1;$$

С учетом того, что $u_1 + u_2 = 1$, получаем:

$$u_1 + 3u_1 = 1; \qquad u_1 = 0{,}25; \qquad u_2 = 0{,}75;$$

\Получаем: $P^{(\infty)} = \begin{pmatrix} 0{,}25 & 0{,}75 \\ 0{,}25 & 0{,}75 \end{pmatrix}$

Лекция 14.

Основные понятия математической статистики. Генеральная совокупность и выборка. Вариационный ряд, статистический ряд. Группированная выборка. Группированный статистический ряд. Полигон частот. Выборочная функция распределения и гистограмма.

Математическая статистика занимается установлением закономерностей, которым подчинены массовые случайные явления, на основе обработки статистических данных, полученных в результате наблюдений. Двумя основными задачами математической статистики являются:

- определение способов сбора и группировки этих статистических данных;
- разработка методов анализа полученных данных в зависимости от целей исследования, к которым относятся:

а) оценка неизвестной вероятности события; оценка неизвестной функции распределения; оценка параметров распределения, вид которого известен; оценка зависимости от других случайных величин и т.д.;

б) проверка статистических гипотез о виде неизвестного распределения или о значениях параметров известного распределения.

Для решения этих задач необходимо выбрать из большой совокупности однородных объектов ограниченное количество объектов, по результатам изучения которых можно сделать прогноз относительно исследуемого признака этих объектов.

Определим основные понятия математической статистики.

Генеральная совокупность – все множество имеющихся объектов.

Выборка – набор объектов, случайно отобранных из генеральной совокупности.

Объем генеральной совокупности *N* и объем выборки *n* – число объектов в рассматриваемой совокупности.

Виды выборки:

Повторная – каждый отобранный объект перед выбором следующего возвращается в генеральную совокупность;

Бесповторная – отобранный объект в генеральную совокупность не возвращается.

Замечание. Для того, чтобы по исследованию выборки можно было сделать выводы о поведении интересующего нас признака генеральной совокупности, нужно, чтобы выборка правильно представляла пропорции генеральной совокупности, то есть была **репрезентативной** (представительной). Учитывая закон больших чисел, можно утверждать, что это условие выполняется, если каждый объект выбран случайно, причем для любого объекта вероятность попасть в выборку одинакова.

Первичная обработка результатов.

Пусть интересующая нас случайная величина X принимает в выборке значение $x_1 - n_1$ раз, $x_2 - n_2$ раз, …, $x_к - n_к$ раз, причем $\sum_{i=1}^{k} n_k = n,$ где n – объем выборки. Тогда наблюдаемые значения случайной величины $x_1, x_2, \ldots, x_к$ называют **вариантами**, а $n_1, n_2, \ldots, n_к$ – **частотами**. Если разделить каждую частоту на объем выборки, то получим **относительные частоты** $w_i = \frac{n_i}{n}$.

Последовательность вариант, записанных в порядке возрастания, называют **вариационным** рядом, а перечень вариант и соответствующих им частот или относительных частот – **статистическим рядом**:

x_i	x_1	x_2	…	x_k
n_i	n_1	n_2	…	n_k
w_i	w_1	w_2	…	w_k

Пример.

При проведении 20 серий из 10 бросков игральной кости число выпадений шести очков оказалось равным 1,1,4,0,1,2,1,2,2,0,5,3,3,1,0,2,2,3,4,1.Составим вариационный ряд: 0,1,2,3,4,5. Статистический ряд для абсолютных и относительных частот имеет вид:

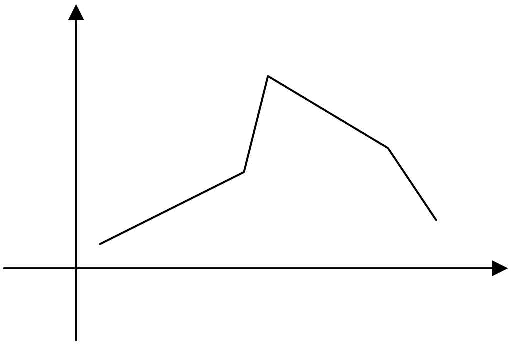
x_i	0	1	2	3	4	5
n_i	3	6	5	3	2	1
w_i	0,15	0,3	0,25	0,15	0,1	0,05

Если исследуется некоторый непрерывный признак, то вариационный ряд может состоять из очень большого количества чисел. В этом случае удобнее использовать **группированную выборку**. Для ее получения интервал, в котором заключены все наблюдаемые значения признака, разбивают на несколько равных частичных интервалов длиной h, а затем находят для каждого частичного интервала n_i – сумму частот вариант, попавших в i-й интервал. Составленная по этим результатам таблица называется **группированным статистическим рядом**:

Номера интервалов	1	2	…	k
Границы интервалов	$(a, a+h)$	$(a+h, a+2h)$	…	$(b-h, b)$
Сумма частот вариант, попавших в интервал	n_1	n_2	…	n_k

Полигон частот. Выборочная функция распределения и гистограмма.

Для наглядного представления о поведении исследуемой случайной величины в выборке можно строить различные графики. Один из них – **полигон частот**: ломаная, отрезки которой соединяют точки с координатами (x_1, n_1), (x_2, n_2),…, (x_k, n_k), где x_i откладываются на оси абсцисс, а n_i – на оси ординат. Если на оси ординат откладывать не абсолютные (n_i), а относительные (w_i) частоты, то получим **полигон относительных частот**



(рис.1). Рис. 1.

По аналогии с функцией распределения случайной величины можно задать некоторую функцию, относительную частоту события $X < x$.

<u>Определение.</u> Выборочной (эмпирической) функцией распределения называют функцию $F^*(x)$, определяющую для каждого значения x относительную частоту события $X < x$. Таким образом,

$$F^*(x) = \frac{n_x}{n},$$

где n_x – число вариант, меньших x, n – объем выборки.

Замечание. В отличие от эмпирической функции распределения, найденной опытным путем, функцию распределения $F(x)$ генеральной совокупности называют *теоретической функцией распределения*. $F(x)$ определяет вероятность события $X < x$, а $F^*(x)$ – его относительную частоту. При достаточно больших n, как следует из теоремы Бернулли, $F^*(x)$ стремится по вероятности к $F(x)$.

Из определения эмпирической функции распределения видно, что ее свойства совпадают со свойствами $F(x)$, а именно:

1) $0 \le F^*(x) \le 1$.

2) $F^*(x)$ – неубывающая функция.

3) Если x_1 – наименьшая варианта, то $F^*(x) = 0$ при $x \le x_1$; если $x_к$ – наибольшая варианта, то $F^*(x) = 1$ при $x > x_к$.

Для непрерывного признака графической иллюстрацией служит **гистограмма**, то есть ступенчатая фигура, состоящая из прямоугольников, основаниями которых служат частичные интервалы длиной h, а высотами – отрезки длиной n_i / h (гистограмма частот) или w_i / h (гистограмма относительных частот). В первом случае площадь гистограммы равна объему выборки, во втором – единице (рис.2).

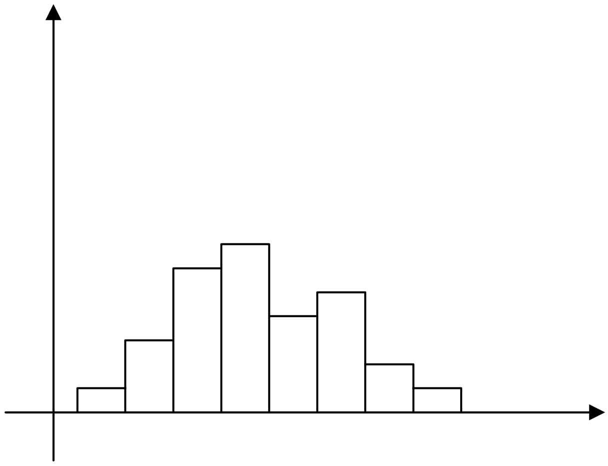


Рис.2.

Лекция 15.

Числовые характеристики статистического распределения: выборочное среднее, оценки дисперсии, оценки моды и медианы, оценки начальных и центральных моментов. Статистическое описание и вычисление оценок параметров двумерного случайного вектора.

Одна из задач математической статистики: по имеющейся выборке оценить значения числовых характеристик исследуемой случайной величины.

Определение. **Выборочным средним** называется среднее арифметическое значений случайной величины, принимаемых в выборке:

$$\overline{x}_B = \frac{x_1 + x_2 + \ldots + x_n}{n} = \frac{n_1 x_1 + n_2 x_2 + \ldots + n_k x_k}{n} = \frac{\sum_{i=1}^{k} n_i x_i}{n},$$

где x_i – варианты, n_i - частоты.

Замечание. Выборочное среднее служит для оценки математического ожидания исследуемой случайной величины. В дальнейшем будет рассмотрен вопрос, насколько точной является такая оценка.

Определение. **Выборочной дисперсией** называется

$$D_B = \frac{\sum_{i=1}^{n}(x_i - \bar{x}_B)^2}{n} = \frac{\sum_{i=1}^{k} n_i (x_i - \bar{x}_B)^2}{n},$$

а **выборочным средним квадратическим отклонением** –

$$\sigma_B = \sqrt{D_B}.$$

Так же, как в теории случайных величин, можно доказать, что справедлива следующая формула для вычисления выборочной дисперсии:

$$D = \overline{x^2} - (\bar{x})^2.$$

Пример 1. Найдем числовые характеристики выборки, заданной статистическим рядом

x_i	2	5	7	8
n_i	3	8	7	2

$$\tilde{o}_{\acute{A}} = \frac{2\cdot 3 + 5\cdot 8 + 7\cdot 7 + 8\cdot 2}{20} = 5{,}55; \quad D_B = \frac{4\cdot 3 + 25\cdot 8 + 49\cdot 7 + 64\cdot 2}{20} - 5{,}55^2 = 3{,}3475;$$

$$\sigma_B = \sqrt{3{,}3475} = 1{,}83.$$

Другими характеристиками вариационного ряда являются:

- **мода M_0** – варианта, имеющая наибольшую частоту (в предыдущем примере $M_0 = 5$).

- **медиана m_e** - варианта, которая делит вариационный ряд на две части, равные по числу вариант. Если число вариант нечетно ($n = 2k + 1$), то $m_e = x_{k+1}$, а при четном $n = 2k$ $m_e = \frac{x_k + x_{k+1}}{2}$. В частности, в примере 1 $m_e = \frac{5+7}{2} = 6.$

Оценки начальных и центральных моментов (так называемые эмпирические моменты) определяются аналогично соответствующим теоретическим моментам:

- начальным эмпирическим моментом порядка *k* называется

$$M_k = \frac{\sum n_i x_i^k}{n}.$$

В частности, $M_1 = \dfrac{\sum n_i x_i}{n} = \bar{x}_B$, то есть начальный эмпирический момент первого порядка равен выборочному среднему.

- **центральным эмпирическим моментом порядка *k*** называется

$$m_k = \frac{\sum n_i (x_i - \bar{x}_B)^k}{n}.$$

В частности, $m_2 = \dfrac{\sum n_i (x_i - \bar{x}_B)^2}{n} = D_B$, то есть центральный эмпирический момент второго порядка равен выборочной дисперсии.

Статистическое описание и вычисление характеристик двумерного случайного вектора.

При статистическом исследовании двумерных случайных величин основной задачей является обычно выявление связи между составляющими. Двумерная выборка представляет собой набор значений случайного вектора: (x_1, y_1), (x_2, y_2), …, (x_n, y_n). Для нее можно определить выборочные средние составляющих: $\bar{x}_B = \dfrac{\sum x_i}{n}$, $\bar{y}_B = \dfrac{\sum y_i}{n}$ и соответствующие выборочные дисперсии и средние квадратические отклонения. Кроме того, можно вычислить **условные средние**: $\bar{y}_x$ - среднее арифметическое наблюдавшихся значений Y, соответствующих $X = x$, и $\bar{x}_y$ - среднее значение наблюдавшихся значений X, соответствующих $Y = y$.

Если существует зависимость между составляющими двумерной случайной величины, она может иметь разный вид: функциональная зависимость, если каждому возможному значению X соответствует одно значение Y, и статистическая, при которой изменение одной величины приводит к изменению распределения другой. Если при этом в результате изменения одной величины меняется среднее значение другой, то статистическую зависимость между ними называют корреляционной.

Лекция 16.

Основные свойства статистических характеристик параметров распределения: несмещенность, состоятельность, эффективность. Несмещенность и состоятельность выборочного среднего как оценки математического ожидания. Смещенность выборочной дисперсии. Пример несмещенной оценки дисперсии. Асимптотически несмещенные оценки. Способы построения оценок: метод наибольшего правдоподобия, метод моментов, метод квантили, метод наименьших квадратов, байесовский подход к получению оценок.

Получив статистические оценки параметров распределения (выборочное среднее, выборочную дисперсию и т.д.), нужно убедиться, что они в достаточной степени служат приближением соответствующих характеристик генеральной совокупности. Определим требования, которые должны при этом выполняться.

Пусть Θ* - статистическая оценка неизвестного параметра Θ теоретического распределения. Извлечем из генеральной совокупности несколько выборок одного и того же объема *n* и вычислим для каждой из них оценку параметра Θ: $\Theta_1^*, \Theta_2^*, ..., \Theta_k^*$. Тогда оценку Θ* можно рассматривать как случайную величину, принимающую возможные значения $\Theta_1^*, \Theta_2^*, ..., \Theta_k^*$. Если математическое ожидание Θ* не равно оцениваемому параметру, мы будем получать при вычислении оценок систематические ошибки одного знака (с избытком, если $M(\Theta^*) > \Theta$, и с недостатком, если $M(\Theta^*) < \Theta$). Следовательно, необходимым условием отсутствия систематических ошибок является требование $M(\Theta^*) = \Theta$.

<u>Определение.</u> Статистическая оценка Θ* называется **несмещенной**, если ее математичес-кое ожидание равно оцениваемому параметру Θ при любом объеме выборки:

$$M(\Theta^*) = \Theta.$$

Смещенной называют оценку, математическое ожидание которой не равно оцениваемому параметру.

Однако несмещенность не является достаточным условием хорошего приближения к истинному значению оцениваемого параметра. Если при этом возможные значения Θ^* могут значительно отклоняться от среднего значения, то есть дисперсия Θ^* велика, то значение, найденное по данным одной выборки, может значительно отличаться от оцениваемого параметра. Следовательно, требуется наложить ограничения на дисперсию.

Определение. Статистическая оценка называется **эффективной**, если она при заданном объеме выборки n имеет наименьшую возможную дисперсию.

При рассмотрении выборок большого объема к статистическим оценкам предъявляется еще и требование состоятельности.

Определение. **Состоятельной** называется статистическая оценка, которая при $n\to\infty$ стремится по вероятности к оцениваемому параметру (если эта оценка несмещенная, то она будет состоятельной, если при $n\to\infty$ ее дисперсия стремится к 0).

Убедимся, что $\bar{x}_B$ представляет собой несмещенную оценку математического ожидания $M(X)$.

Будем рассматривать $\bar{x}_B$ как случайную величину, а $x_1, x_2,\dots, x_n$, то есть значения исследуемой случайной величины, составляющие выборку, – как независимые, одинаково распределенные случайные величины $X_1, X_2,\dots, X_n$, имеющие математическое ожидание a. Из свойств математического ожидания следует, что

$$M(\bar{X}_B) = M\left(\frac{X_1 + X_2 + \ldots + X_n}{n}\right) = a.$$

Но, поскольку каждая из величин $X_1, X_2,\dots, X_n$ имеет такое же распределение, что и генеральная совокупность, $a = M(X)$, то есть $M(\bar{X}_B) = M(X)$, что и требовалось доказать. Выборочное среднее является не только несмещенной, но и состоятельной оценкой математического ожидания. Если предположить, что $X_1, X_2, \ldots, X_n$ имеют ограниченные дисперсии, то из теоремы Чебышева следует, что их среднее арифметическое, то есть $\overline{X}_B$, при увеличении n стремится по вероятности к математическому ожиданию a каждой их величин, то есть к $M(X)$. Следовательно, выборочное среднее есть состоятельная оценка математического ожидания.

В отличие от выборочного среднего, выборочная дисперсия является смещенной оценкой дисперсии генеральной совокупности. Можно доказать, что

$$M(D_B) = \frac{n-1}{n} D_Г,$$

где $D_Г$ – истинное значение дисперсии генеральной совокупности. Можно предложить другую оценку дисперсии – **исправленную дисперсию s^2**, вычисляемую по формуле

$$s^2 = \frac{n}{n-1} D_B = \frac{\sum_{i=1}^{k} n_i (x_i - \overline{x}_B)^2}{n-1}.$$

Такая оценка будет являться несмещенной. Ей соответствует **исправленное среднее квадратическое отклонение**

$$s = \sqrt{s^2} = \sqrt{\frac{\sum_{i=1}^{k} n_i (x_i - \overline{x}_B)^2}{n-1}}. \qquad ($$

<u>**Определение.**</u> Оценка некоторого признака называется **асимптотически несмещенной**, если для выборки $x_1, x_2, \ldots, x_n$

$$\lim_{n \to \infty} \frac{x_1 + x_2 + \ldots + x_n}{n} = X,$$

где X – истинное значение исследуемой величины.

Способы построения оценок.

1. Метод наибольшего правдоподобия.

Пусть X – дискретная случайная величина, которая в результате n испытаний приняла значения $x_1, x_2, \ldots, x_n$. Предположим, что нам известен закон распределения этой величины, определяемый параметром Θ, но неизвестно численное значение этого параметра. Найдем его точечную оценку.

Пусть $p(x_i, \Theta)$ – вероятность того, что в результате испытания величина X примет значение x_i. Назовем **функцией правдоподобия** дискретной случайной величины X функцию аргумента Θ, определяемую по формуле:

$$L\,(x_1, x_2, \ldots, x_n; \Theta) = p(x_1,\Theta)p(x_2,\Theta)\ldots p(x_n,\Theta).$$

Тогда в качестве точечной оценки параметра Θ принимают такое его значение $\Theta^* = \Theta(x_1, x_2, \ldots, x_n)$, при котором функция правдоподобия достигает максимума. Оценку Θ^* называют **оценкой наибольшего правдоподобия**.

Поскольку функции L и $\ln L$ достигают максимума при одном и том же значении Θ, удобнее искать максимум $\ln L$ – **логарифмической функции правдоподобия**. Для этого нужно:

1) найти производную $\dfrac{d\ln L}{d\Theta}$;

2) приравнять ее нулю (получим так называемое *уравнение правдоподобия*) и найти критическую точку;

3) найти вторую производную $\dfrac{d^2\ln L}{d\Theta^2}$; если она отрицательна в критической точке, то это – точка максимума.

Достоинства метода наибольшего правдоподобия: полученные оценки состоятельны (хотя могут быть смещенными), распределены асимптотически нормально при больших значениях n и имеют наименьшую дисперсию по сравнению с другими асимптотически нормальными оценками; если для оцениваемого параметра Θ существует эффективная оценка Θ^*, то уравнение правдоподобия имеет единственное решение Θ^*; метод наиболее полно использует данные выборки и поэтому особенно полезен в случае малых выборок.

Недостаток метода наибольшего правдоподобия: сложность вычислений.

Для непрерывной случайной величины с известным видом плотности распределения $f(x)$ и неизвестным параметром Θ функция правдоподобия имеет вид:

$$L\ (x_1, x_2, \ldots, x_n; \Theta) = f(x_1,\Theta)f(x_2,\Theta)\ldots f(x_n,\Theta).$$

Оценка наибольшего правдоподобия неизвестного параметра проводится так же, как для дискретной случайной величины.

2. Метод моментов.

Метод моментов основан на том, что начальные и центральные эмпирические моменты являются состоятельными оценками соответственно начальных и центральных теоретических моментов, поэтому можно приравнять теоретические моменты соответствующим эмпирическим моментам того же порядка.

Если задан вид плотности распределения $f(x, \Theta)$, определяемой одним неизвестным параметром Θ, то для оценки этого параметра достаточно иметь одно уравнение. Например, можно приравнять начальные моменты первого порядка:

$$\overline{x}_B = M(X) = \int_{-\infty}^{\infty} xf(x;\Theta)dx = \varphi(\Theta),$$

получив тем самым уравнение для определения Θ. Его решение Θ^* будет точечной оценкой параметра, которая является функцией от выборочного среднего и, следовательно, и от вариант выборки:

$$\Theta = \psi\ (x_1, x_2, \ldots, x_n).$$

Если известный вид плотности распределения $f(x, \Theta_1, \Theta_2)$ определяется двумя неизвестными параметрами Θ_1 и Θ_2, то требуется составить два уравнения, например

$$\nu_1 = M_1,\ \ \mu_2 = m_2.$$

Отсюда $\begin{cases} M(X) = \bar{x}_B \\ D(X) = D_B \end{cases}$ - система двух уравнений с двумя неизвестными Θ_1 и Θ_2. Ее решениями будут точечные оценки Θ_1* и Θ_2* - функции вариант выборки:

$$\Theta_1 = \psi_1 (x_1, x_2, \ldots, x_n),$$

$$\Theta_2 = \psi_2(x_1, x_2, \ldots, x_n).$$

3. Метод наименьших квадратов.

Если требуется оценить зависимость величин y и x, причем известен вид связывающей их функции, но неизвестны значения входящих в нее коэффициентов, их величины можно оценить по имеющейся выборке с помощью метода наименьших квадратов. Для этого функция $y = \varphi (x)$ выбирается так, чтобы сумма квадратов отклонений наблюдаемых значений $y_1, y_2,\ldots, y_n$ от $\varphi(x_i)$ была минимальной:

$$\sum_{i=1}^{n}(y_i - \varphi(x_i))^2 = \min.$$

При этом требуется найти стационарную точку функции $\varphi(x;\ a, b, c\ldots)$, то есть решить систему:

$$\begin{cases} \sum\limits_{i=1}^{n}(y_i - \varphi(x_i; a,b,c...))\left(\dfrac{\partial\varphi}{\partial a}\right)_i = 0 \\ \sum\limits_{i=1}^{n}(y_i - \varphi(x_i; a,b,c...))\left(\dfrac{\partial\varphi}{\partial b}\right)_i = 0 \\ \sum\limits_{i=1}^{n}(y_i - \varphi(x_i; a,b,c...))\left(\dfrac{\partial\varphi}{\partial c}\right)_i = 0 \\ \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \end{cases}$$

(решение, конечно, возможно только в случае, когда известен конкретный вид функции φ).

Рассмотрим в качестве примера подбор параметров линейной функции методом наименьших квадратов.

Для того, чтобы оценить параметры a и b в функции $y = ax + b$, найдем

$$\left(\frac{\partial\varphi}{\partial a}\right)_i = x_i; \left(\frac{\partial\varphi}{\partial b}\right)_i = 1.$$

Тогда $\begin{cases} \sum_{i=1}^{n}(y_i - (ax_i + b))x_i = 0 \\ \sum_{i=1}^{n}(y_i - (ax_i + b)) = 0 \end{cases}$.

Отсюда $\begin{cases} \sum_{i=1}^{n}x_i y_i - a\sum_{i=1}^{n}x_i^2 - b\sum_{i=1}^{n}x_i = 0 \\ \sum_{i=1}^{n}y_i - a\sum_{i=1}^{n}x_i - bn = 0 \end{cases}$.

Разделив оба полученных уравнения на n и вспомнив определения эмпирических моментов, можно получить выражения для a и b в виде:

$a = \frac{(K_{xy})_B}{(D_x)_B}$, $b = \bar{y}_B - \frac{(K_{xy})_B}{(D_x)_B}\bar{x}_B$.

Следовательно, связь между x и y можно задать в виде:

$$y - \bar{y}_B = \frac{(K_{xy})_B}{(D_x)_B}(x - \bar{x}_B).$$

4. Байесовский подход к получению оценок.

Пусть (Y, X) – случайный вектор, для которого известна плотность $p(y|x)$ условного распределения Y при каждом значении $X = x$. Если в результате эксперимента получены лишь значения Y, а соответствующие значения X неизвестны, то для оценки некоторой заданной функции $\varphi(x)$ в качестве ее приближенного значения предлагается искать условное математическое ожидание $M(\varphi(x)|Y)$, вычисляемое по формуле:

$$\psi(Y) = \frac{\int \varphi(x) p(Y \mid x) p(x) d\mu(x)}{q(Y)},$$

где $q(y) = \int p(y \mid x) p(x) d\mu(x)$, $p(x)$ – плотность безусловного распределения X, $q(y)$ – плотность безусловного распределения Y. Задача может быть решена только тогда, когда известна $p(x)$. Иногда, однако, удается построить состоятельную оценку для $q(y)$, зависящую только от полученных в выборке значений Y.

Лекция 17.

Интервальное оценивание неизвестных параметров. Точность оценки, доверительная вероятность (надежность), доверительный интервал. Построение доверительных интервалов для оценки математического ожидания нормального распределения при известной и при неизвестной дисперсии. Доверительные интервалы для оценки среднего квадратического отклонения нормального распределения.

При выборке малого объема точечная оценка может значительно отличаться от оцениваемого параметра, что приводит к грубым ошибкам. Поэтому в таком случае лучше пользоваться *интервальными оценками*, то есть указывать интервал, в который с заданной вероятностью попадает истинное значение оцениваемого параметра. Разумеется, чем меньше длина этого интервала, тем точнее оценка параметра. Поэтому, если для оценки Θ^* некоторого параметра Θ справедливо неравенство $| \Theta^* - \Theta | < \delta$, число $\delta > 0$ характеризует **точность оценки** (чем меньше δ, тем точнее оценка). Но статистические методы позволяют говорить только о том, что это неравенство выполняется с некоторой вероятностью.

Определение. **Надежностью (доверительной вероятностью)** оценки Θ^* параметра Θ называется вероятность γ того, что выполняется неравенство $| \Theta^* - \Theta | < \delta$. Если заменить это неравенство двойным неравенством $-\delta < \Theta^* - \Theta < \delta$, то получим:

$$p (\Theta^* - \delta < \Theta < \Theta^* + \delta) = \gamma.$$

Таким образом, γ есть вероятность того, что Θ попадает в интервал ($\Theta^* - \delta, \Theta^* + \delta$).

<u>Определение.</u> Доверительным называется интервал, в который попадает неизвестный параметр с заданной надежностью γ.

Построение доверительных интервалов.

1. Доверительный интервал для оценки математического ожидания нормального распределения при известной дисперсии.

Пусть исследуемая случайная величина X распределена по нормальному закону с известным средним квадратическим σ, и требуется по значению выборочного среднего $\bar{x}_B$ оценить ее математическое ожидание a. Будем рассматривать выборочное среднее $\bar{x}_B$ как случайную величину $\bar{X}$, а значения вариант выборки $x_1, x_2,\dots, x_n$ как одинаково распределенные независимые случайные величины $X_1, X_2,\dots, X_n$, каждая из которых имеет математическое ожидание a и среднее квадратическое отклонение σ. При этом $M(\bar{X}) = a$, $\sigma(\bar{X}) = \frac{\sigma}{\sqrt{n}}$ (используем свойства математического ожидания и дисперсии суммы независимых случайных величин). Оценим вероятность выполнения неравенства $|\bar{X} - a| < \delta$. Применим формулу для вероятности попадания нормально распределенной случайной величины в заданный интервал:

$p\,(|X - a| < \delta) = 2\Phi\left(\frac{\delta}{\sigma}\right)$. Тогда , с учетом того, что $\sigma(\bar{X}) = \frac{\sigma}{\sqrt{n}}$, $p\,(|\bar{X} - a| < \delta) =$ $2\Phi\left(\frac{\delta\sqrt{n}}{\sigma}\right)$=$2\Phi(t)$, где $t = \frac{\delta\sqrt{n}}{\sigma}$. Отсюда $\delta = \frac{t\sigma}{\sqrt{n}}$, и предыдущее равенство можно переписать так:

$$p\left(\bar{x}_B - \frac{t\sigma}{\sqrt{n}} < a < \bar{x}_B + \frac{t\sigma}{\sqrt{n}}\right) = 2\Phi(t) = \gamma\,.$$

Итак, значение математического ожидания a с вероятностью (надежностью) γ попадает в интервал $\left(\bar{x}_B - \frac{t\sigma}{\sqrt{n}}; \bar{x}_B + \frac{t\sigma}{\sqrt{n}}\right)$, где значение t

определяется из таблиц для функции Лапласа так, чтобы выполнялось равенство $2\Phi(t) = \gamma$.

Пример. Найдем доверительный интервал для математического ожидания нормально распределенной случайной величины, если объем выборки $n = 49$, $\bar{x}_B = 2{,}8$, $\sigma = 1{,}4$, а доверительная вероятность $\gamma = 0{,}9$. Определим t, при котором $\Phi(t) = 0{,}9:2 = 0{,}45$: $t = 1{,}645$. Тогда $2{,}8 - \frac{1{,}645 \cdot 1{,}4}{\sqrt{49}} < a < 2{,}8 + \frac{1{,}645 \cdot 1{,}4}{\sqrt{14}}$, или $2{,}471 < a < 3{,}129$. Найден доверительный интервал, в который попадает a с надежностью 0,9.

2. Доверительный интервал для оценки математического ожидания нормального распределения при неизвестной дисперсии.

Если известно, что исследуемая случайная величина X распределена по нормальному закону с неизвестным средним квадратическим отклонением, то для поиска доверительного интервала для ее математического ожидания построим новую случайную величину

$$T = \frac{\bar{x}_B - a}{\frac{s}{\sqrt{n}}},$$

где $\bar{x}_B$ - выборочное среднее, s – исправленная дисперсия, n – объем выборки. Эта случайная величина, возможные значения которой будем обозначать t, имеет распределение Стьюдента (см. лекцию 12) с $k = n - 1$ степенями свободы.

Поскольку плотность распределения Стьюдента $s(t,n) = B_n \left(1 + \frac{t^2}{n-1}\right)^{-\frac{n}{2}}$,

где $B_n = \dfrac{\Gamma\left(\frac{n}{2}\right)}{\sqrt{\pi(n-1)}\,\Gamma\left(\frac{n-1}{2}\right)}$, явным образом не зависит от a и σ, можно задать вероятность ее попадания в некоторый интервал ($-\,t_\gamma$, t_γ), учитывая четность плотности распределения, следующим образом:

$$p\left(\left|\frac{\overline{x}_B - a}{\frac{s}{\sqrt{n}}}\right| < t_\gamma\right) = 2\int_0^{t_\gamma} s(t,n)dt = \gamma \ .$$

Отсюда получаем:

$$p\left(\overline{x}_B - \frac{t_\gamma s}{\sqrt{n}} < a < \overline{x}_B + \frac{t_\gamma s}{\sqrt{n}}\right) = \gamma.$$

Таким образом, получен доверительный интервал для a, где t_γ можно найти по соответствующей таблице при заданных n и γ.

Пример. Пусть объем выборки $n = 25$, $\overline{x}_B = 3$, $s = 1{,}5$. Найдем доверительный интервал для a при $\gamma = 0{,}99$. Из таблицы находим, что t_γ ($n = 25$, $\gamma = 0{,}99$) = 2,797. Тогда $3 - \frac{2{,}797 \cdot 1{,}5}{\sqrt{25}} < a < 3 + \frac{2{,}797 \cdot 1{,}5}{\sqrt{25}}$, или $2{,}161 < a < 3{,}839$ – доверительный интервал, в который попадает a с вероятностью 0,99.

3. Доверительные интервалы для оценки среднего квадратического отклонения нормального распределения.

Будем искать для среднего квадратического отклонения нормально распределенной случайной величины доверительный интервал вида $(s - \delta, s + \delta)$, где s – исправленное выборочное среднее квадратическое отклонение, а для δ выполняется условие: $p\ (\ |\sigma - s| < \delta\ ) = \gamma$.

Запишем это неравенство в виде: $s\left(1 - \frac{\delta}{s}\right) < \sigma < s\left(1 + \frac{\delta}{s}\right)$ или, обозначив $q = \frac{\delta}{s}$,

$$s(1-q) < \sigma < s(1+q).$$

Рассмотрим случайную величину χ, определяемую по формуле

$$\chi = \frac{s}{\sigma}\sqrt{n-1}\ ,$$

которая распределена по закону «хи-квадрат» с n-1 степенями свободы (см. лекцию 12). Плотность ее распределения не зависит от оцениваемого параметра σ, а зависит только от объема выборки n. Преобразуем неравенство (18.4) так, чтобы оно приняло вид $\chi_1 < \chi < \chi_2$. Вероятность выполнения этого неравенства равна доверительной вероятности γ, следовательно, $\int_{\chi_1}^{\chi_2} R(\chi, n)d\chi = \gamma.$ Предположим, что $q < 1$, тогда неравенство (18.4) можно записать так:

$$\frac{1}{s(1+q)} < \frac{1}{\sigma} < \frac{1}{s(1-q)},$$

или, после умножения на $s\sqrt{n-1}$, $\frac{\sqrt{n-1}}{1+q} < \frac{s\sqrt{n-1}}{\sigma} < \frac{\sqrt{n-1}}{1-q}$. Следовательно, $\frac{\sqrt{n-1}}{1+q} < \chi < \frac{\sqrt{n-1}}{1-q}$. Тогда $\int_{\frac{\sqrt{n-1}}{1+q}}^{\frac{\sqrt{n-1}}{1-q}} R(\chi, n)d\chi = \gamma.$ Существуют таблицы для распределения «хи-квадрат», из которых можно найти q по заданным n и γ, не решая этого уравнения. Таким образом, вычислив по выборке значение s и определив по таблице значение q, можно найти доверительный интервал (18.4), в который значение σ попадает с заданной вероятностью γ.

Замечание. Если $q > 1$, то с учетом условия $\sigma > 0$ доверительный интервал для σ будет иметь границы

$$0 < \sigma < s(1+q).$$

Пример. Пусть $n = 20$, $s = 1{,}3$. Найдем доверительный интервал для σ при заданной надежности $\gamma = 0{,}95$. Из соответствующей таблицы находим q ($n = 20$, $\gamma = 0{,}95$) = 0,37. Следовательно, границы доверительного интервала: 1,3(1-0,37) = 0,819 и 1,3(1+0,37) = 1,781. Итак, $0{,}819 < \sigma < 1{,}781$ с вероятностью 0,95.

Лекция 18.

Статистическая проверка статистических гипотез. Общие принципы проверки гипотез. Понятия статистической гипотезы (простой и сложной), нулевой и конкурирующей гипотезы, ошибок первого и второго рода, уровня значимости, статистического критерия, критической области, области принятия гипотезы. Наблюдаемое значение критерия.

Критические точки. Мощность критерия. Критерии для проверки гипотез о вероятности события, о математическом ожидании, о сравнении двух дисперсий.

<u>Определение.</u> Статистической гипотезой называют гипотезу о виде неизвестного распределения генеральной совокупности или о параметрах известных распределений.

<u>Определение.</u> Нулевой (основной) называют выдвинутую гипотезу H_0. **Конкурирующей (альтернативной)** называют гипотезу H_1, которая противоречит нулевой.

Пример. Пусть H_0 заключается в том, что математическое ожидание генеральной совокупности $a = 3$. Тогда возможные варианты H_1: а) $a \neq 3$; б) $a > 3$; в) $a < 3$.

<u>Определение.</u> Простой называют гипотезу, содержащую только одно предположение, **сложной** – гипотезу, состоящую из конечного или бесконечного числа простых гипотез.

Пример. Для показательного распределения гипотеза H_0: $\lambda = 2$ – простая, H_0: $\lambda > 2$ – сложная, состоящая из бесконечного числа простых (вида $\lambda = c$, где c – любое число, большее 2).

В результате проверки правильности выдвинутой нулевой гипотезы (такая проверка называется **статистической**, так как производится с применением методов математической статистики) возможны ошибки двух видов: **ошибка первого рода**, состоящая в том, что будет отвергнута правильная нулевая гипотеза, и **ошибка второго рода**, заключающаяся в том, что будет принята неверная гипотеза.

Замечание. Какая из ошибок является на практике более опасной, зависит от конкретной задачи. Например, если проверяется правильность выбора метода лечения больного, то ошибка первого рода означает отказ от правильной методики, что может замедлить лечение, а ошибка второго рода (применение неправильной методики) чревата ухудшением состояния больного и является более опасной.

Определение. Вероятность ошибки первого рода называется **уровнем значимости α.**

Основной прием проверки статистических гипотез заключается в том, что по имеющейся выборке вычисляется значение некоторой случайной величины, имеющей известный закон распределения.

Определение. **Статистическим критерием** называется случайная величина K с известным законом распределения, служащая для проверки нулевой гипотезы.

Определение. **Критической областью** называют область значений критерия, при которых нулевую гипотезу отвергают, **областью принятия гипотезы** – область значений критерия, при которых гипотезу принимают.

Итак, процесс проверки гипотезы состоит из следующих этапов:

1) выбирается статистический критерий K;
2) вычисляется его наблюдаемое значение $K_{набл}$ по имеющейся выборке;
3) поскольку закон распределения K известен, определяется (по известному уровню значимости α) **критическое значение** $k_{кр}$, разделяющее критическую область и область принятия гипотезы (например, если

$p(K > k_{кр}) = \alpha$, то справа от $k_{кр}$ распо-лагается критическая область, а слева – область принятия гипотезы);

4) если вычисленное значение $K_{набл}$ попадает в область принятия гипотезы, то нулевая гипотеза принимается, если в критическую область – нулевая гипотеза отвергается.

Различают разные виды критических областей:

- **правостороннюю** критическую область, определяемую неравенством $K > k_{кр}$ ($k_{кр} > 0$);
- **левостороннюю** критическую область, определяемую неравенством $K < k_{кр}$ ($k_{кр} < 0$);

- **двустороннюю** критическую область, определяемую неравенствами $K < k_1, K > k_2 \quad (k_2 > k_1)$.

Определение. **Мощностью критерия** называют вероятность попадания критерия в критическую область при условии, что верна конкурирующая гипотеза.

Если обозначить вероятность ошибки второго рода (принятия неправильной нулевой гипотезы) β, то мощность критерия равна 1 – β. Следовательно, чем больше мощность критерия, тем меньше вероятность совершить ошибку второго рода. Поэтому после выбора уровня значимости следует строить критическую область так, чтобы мощность критерия была максимальной.

Критерий для проверки гипотезы о вероятности события.

Пусть проведено n независимых испытаний (n – достаточно большое число), в каждом из которых некоторое событие A появляется с одной и той же, но неизвестной вероятностью p, и найдена относительная частота $\frac{m}{n}$ появлений A в этой серии испытаний. Проверим при заданном уровне значимости α нулевую гипотезу H_0, состоящую в том, что вероятность p равна некоторому значению p_0.

Примем в качестве статистического критерия случайную величину

$$U = \frac{\left(\frac{M}{n} - p_0\right)\sqrt{n}}{\sqrt{p_0 q_0}},$$

имеющую нормальное распределение с параметрами $M(U) = 0$, $\sigma(U) = 1$ (то есть нормированную). Здесь $q_0 = 1 - p_0$. Вывод о нормальном распределении критерия следует из теоремы Лапласа (при достаточно большом n относительную частоту можно приближенно считать нормально распределенной с математическим ожиданием p и средним квадратическим отклонением $\sqrt{\frac{pq}{n}}$).

Критическая область строится в зависимости от вида конкурирующей гипотезы.

1) Если H_0: $p = p_0$, а H_1: $p \neq p_0$, то критическую область нужно построить так, чтобы вероятность попадания критерия в эту область равнялась заданному уровню значимости α. При этом наибольшая мощность критерия достигается тогда, когда критическая область состоит из двух интервалов, вероятность попадания в каждый из которых равна $\frac{\alpha}{2}$. Поскольку U симметрична относительно оси Оy, вероятность ее попадания в интервалы (-∞; 0) и (0; +∞) равна 0,5, следовательно, критическая область тоже должна быть симметрична относительно Оy. Поэтому $u_{кр}$ определяется по таблице значений функции Лапласа из условия $\Phi(u_{кр}) = \frac{1-\alpha}{2}$, а критическая область имеет вид $(-\infty; -u_{кр}) \cup (u_{кр}; +\infty)$.

Замечание. Предполагается, что используется таблица значений функции Лапласа, заданной в виде $\Phi(x) = \int_0^x e^{-\frac{t^2}{2}} dt$, где нижний предел интегрирования равен 0, а не -∞. Функция Лапласа, заданная таким образом, является нечетной, а ее значения на 0,5 меньше, чем значения стандартной функции $\Phi(x)$ (см. лекцию 6).

Далее нужно вычислить наблюдаемое значение критерия:

$$U_{набл} = \frac{\left(\frac{m}{n} - p_0\right)\sqrt{n}}{\sqrt{p_0 q_0}}.$$

Если $|U_{набл}| < u_{кр}$, то нулевая гипотеза принимается.

Если $|U_{набл}| > u_{кр}$, то нулевая гипотеза отвергается.

2) Если конкурирующая гипотеза H_1: $p > p_0$, то критическая область определяется неравенством $U > u_{кр}$, то есть является правосторонней, причем $p(U > u_{кр}) = \alpha$. Тогда $p(0 < U < u_{кр}) = \frac{1}{2} - \alpha = \frac{1-2\alpha}{2}$. Следовательно, $u_{кр}$ можно найти по таблице значений функции Лапласа из условия, что $\Phi(u_{кр}) = \frac{1-2\alpha}{2}$. Вычислим наблюдаемое значение критерия по формуле (19.2).

Если $U_{набл} < u_{кр}$, то нулевая гипотеза принимается.

Если $U_{набл} > u_{кр}$, то нулевая гипотеза отвергается.

3) Для конкурирующей гипотезы H_1: $p < p_0$ критическая область является левосторонней и задается неравенством $U <- u_{кр}$, где $u_{кр}$ вычисляется так же, как в предыдущем случае.

Если $U_{набл} > - u_{кр}$, то нулевая гипотеза принимается.

Если $U_{набл} < - u_{кр}$, то нулевая гипотеза отвергается.

Пример. Пусть проведено 50 независимых испытаний, и относительная частота появления события A оказалась равной 0,12. Проверим при уровне значимости $\alpha = 0{,}01$ нулевую гипотезу H_0: $p = 0{,}1$ при конкурирующей гипотезе H_1: $p > 0{,}1$. Найдем $U_{набл} = \frac{(0{,}12-0{,}1)\sqrt{50}}{\sqrt{0{,}1\cdot 0{,}9}} = 0{,}471.$ Критическая область является правосторонней, а $u_{кр}$ находим из равенства

$$\Phi(u_{кр}) = \frac{1-2\cdot 0{,}01}{2} = 0{,}49.$$

Из таблицы значений функции Лапласа определяем $u_{кр} = 2{,}33$.

Итак, $U_{набл} < u_{кр}$, и гипотеза о том, что $p = 0{,}1$, принимается.

Критерий для проверки гипотезы о математическом ожидании.

Пусть генеральная совокупность X имеет нормальное распределение, и требуется проверить предположение о том, что ее математическое ожидание равно некоторому числу a_0. Рассмотрим две возможности.

1) Известна дисперсия σ^2 генеральной совокупности. Тогда по выборке объема n найдем выборочное среднее $\bar{x}_в$ и проверим нулевую гипотезу

$$H_0: M(X) = a_0.$$

Учитывая, что выборочное среднее $\bar{X}$ является несмещенной оценкой $M(X)$, то есть $M(\bar{X}) = M(X)$, можно записать нулевую гипотезу так: $M(\bar{X}) = a_0$. Для ее проверки выберем критерий

$$U = \frac{\bar{X} - a_0}{\sigma(\bar{X})} = \frac{(\bar{X} - a_0)\sqrt{n}}{\sigma}.$$

Это случайная величина, имеющая нормальное распределение, причем, если нулевая гипотеза справедлива, то $M(U) = 0$, $\sigma(U) = 1$.

Выберем критическую область в зависимости от вида конкурирующей гипотезы:

- если H_1: $M(\bar{X}) \neq a_0$, то $u_{кр}$: $\Phi(u_{кр}) = \frac{1-\alpha}{2}$, критическая область двусторонняя, $U_{набл} = \frac{(\bar{x} - a_0)\sqrt{n}}{\sigma}$, и, если $|U_{набл}| < u_{кр}$, то нулевая гипотеза принимается; если $|U_{набл}| > u_{кр}$, то нулевая гипотеза отвергается.

- если H_1: $M(\bar{X}) > a_0$, то $u_{кр}$: $\Phi(u_{кр}) = \frac{1-2\alpha}{2}$, критическая область правосторонняя, и, если $U_{набл} < u_{кр}$, то нулевая гипотеза принимается; если $U_{набл} > u_{кр}$, то нулевая гипотеза отвергается.

- если H_1: $M(\bar{X}) < a_0$, то $u_{кр}$: $\Phi(u_{кр}) = \frac{1-2\alpha}{2}$, критическая область левосторонняя, и, если $U_{набл} > - u_{кр}$, то нулевая гипотеза принимается; если $U_{набл} < - u_{кр}$, то нулевая гипотеза отвергается.

2) Дисперсия генеральной совокупности неизвестна.

В этом случае выберем в качестве критерия случайную величину

$$T = \frac{(\bar{X} - a_0)\sqrt{n}}{S},$$

где S – исправленное среднее квадратическое отклонение. Такая случайная величина имеет распределение Стьюдента с $k = n – 1$ степенями свободы.

Рассмотрим те же, что и в предыдущем случае, конкурирующие гипотезы и соответствующие им критические области. Предварительно вычислим наблюдаемое значение критерия:

$$T_{набл} = \frac{(\bar{x}_B - a_0)\sqrt{n}}{S}.$$

- если H_1: $M(\bar{X}) \neq a_0$, то критическая точка $t_{двуст.кр.}$ находится по таблице критических точек распределения Стьюдента по известным α и $k = n - 1$.
Если $|T_{набл}| < t_{двуст.кр.}$, то нулевая гипотеза принимается.
Если $|T_{набл}| > t_{двуст.кр.}$, то нулевая гипотеза отвергается.
- если H_1: $M(\bar{X}) > a_0$, то по соответствующей таблице находят $t_{правост.кр.}(\alpha, k)$ – критическую точку правосторонней критической области. Нулевая гипотеза принимается, если $T_{набл} < t_{правост.кр.}$.
- при конкурирующей гипотезе H_1: $M(\bar{X}) < a_0$ критическая область является левосторон-ней, и нулевая гипотеза принимается при условии $T_{набл} > - t_{правост.кр.}$. Если $T_{набл} < - t_{правост.кр.}$, нулевую гипотезу отвергают.

Критерий для проверки гипотезы о сравнении двух дисперсий.

Пусть имеются две нормально распределенные генеральные совокупности X и Y. Из них извлечены независимые выборки объемов соответственно n_1 и n_2, по которым вычислены исправленные выборочные дисперсии s_X^2 и s_Y^2. Требуется при заданном уровне значимости α проверить нулевую гипотезу H_0: $D(X) = D(Y)$ о равенстве дисперсий рассматриваемых генеральных совокупностей. Учитывая несмещенность исправленных выборочных дисперсий, можно записать нулевую гипотезу так:

$$H_0: M(s_X^2) = M(s_Y^2).$$

Замечание. Конечно, исправленные дисперсии, вычисленные по выборкам, обычно оказываются различными. При проверке гипотезы выясняется, является ли это различие незначимым и обусловленным случайными причинами (в случае принятия нулевой гипотезы) или оно является следствием того, что сами генеральные дисперсии различны.

В качестве критерия примем случайную величину

$$F = \frac{S_\sigma^2}{S_M^2} \text{ -}$$

- отношение большей выборочной дисперсии к меньшей. Она имеет распределение Фишера-Снедекора со степенями свободы $k_1 = n_1 - 1$ и $k_2 = n_2 - 1$, где n_1 – объем выборки, по которой вычислена большая исправленная дисперсия, а n_2 – объем второй выборки. Рассмотрим два вида конкурирующих гипотез:

- пусть H_1: $D(X) > D(Y)$. Наблюдаемым значением критерия будет отношение большей из исправленных дисперсий к меньшей: $F_{набл} = \frac{s_\sigma^2}{s_M^2}$. По таблице критических точек распределения Фишера-Снедекора можно найти критическую точку $F_{набл}(\alpha; k_1; k_2)$. При $F_{набл} < F_{кр}$ нулевая гипотеза принимается, при $F_{набл} > F_{кр}$ отвергается.

- если H_1: $D(X) \neq D(Y)$, то критическая область является двусторонней и определяется неравенствами $F < F_1$, $F > F_2$, где $p(F < F_1) = p(F > F_2) = \alpha/2$. При этом достаточно найти правую критическую точку $F_2 = F_{кр}\left(\frac{\alpha}{2}, k_1, k_2\right)$. Тогда при $F_{набл} < F_{кр}$ нулевая гипотеза принимается, при $F_{набл} > F_{кр}$ отвергается.

Лекция 18.

Критерий Пирсона для проверки гипотезы о виде закона распределения случайной величины. Проверка гипотез о нормальном, показательном и равномерном распределениях по критерию Пирсона. Критерий Колмогорова. Приближенный метод проверки нормальности распределения, связанный с оценками коэффициентов асимметрии и эксцесса.

В предыдущей лекции рассматривались гипотезы, в которых закон распределения генеральной совокупности предполагался известным. Теперь займемся проверкой гипотез о предполагаемом законе неизвестного распределения, то есть будем проверять нулевую гипотезу о том, что генеральная совокупность распределена по некоторому известному закону. Обычно статистические критерии для проверки таких гипотез называются **критериями согласия**.

Критерий Пирсона.

Достоинством критерия Пирсона является его универсальность: с его помощью можно проверять гипотезы о различных законах распределения.

1. Проверка гипотезы о нормальном распределении.

Пусть получена выборка достаточно большого объема n с большим количеством различных значений вариант. Для удобства ее обработки разделим интервал от наименьшего до наибольшего из значений вариант на s равных частей и будем считать, что значения вариант, попавших в каждый интервал, приближенно равны числу, задающему середину интервала. Подсчитав число вариант, попавших в каждый интервал, составим так называемую сгруппированную выборку:

варианты……….$x_1 \quad x_2 \quad \ldots \quad x_s$

частоты…………$n_1 \quad n_2 \quad \ldots \quad n_s$,

где x_i – значения середин интервалов, а n_i – число вариант, попавших в i-й интервал (эмпирические частоты).

По полученным данным можно вычислить выборочное среднее $\bar{x}_B$ и выборочное среднее квадратическое отклонение σ_B. Проверим предположение, что генеральная совокупность распределена по нормальному закону с параметрами $M(X) = \bar{x}_B$, $D(X) = \sigma_B^2$. Тогда можно найти количество чисел из выборки объема n, которое должно оказаться в каждом интервале при этом предположении (то есть теоретические частоты). Для этого по таблице значений функции Лапласа найдем вероятность попадания в i-й интервал:

$$p_i = \Phi\left(\frac{b_i - \overline{x}_B}{\sigma_B}\right) - \Phi\left(\frac{a_i - \overline{x}_B}{\sigma_B}\right),$$

где a_i и b_i - границы i-го интервала. Умножив полученные вероятности на объем выборки п, найдем теоретические частоты: $n_i = n \cdot p_i$. Наша цель – сравнить эмпирические и теоретические частоты, которые, конечно, отличаются друг от друга, и выяснить, являются ли эти различия несущественными, не опровергающими гипотезу о нормальном распределении исследуемой случайной величины, или они настолько велики, что противоречат этой гипотезе. Для этого используется критерий в виде случайной величины

$$\chi^2 = \sum_{i=1}^{s} \frac{(n_i - n_i')^2}{n_i'} .$$

Смысл ее очевиден: суммируются части, которые квадраты отклонений эмпирических частот от теоретических составляют от соответствующих теоретических частот. Можно доказать, что вне зависимости от реального закона распределения генеральной совокупности закон распределения случайной величины (20.1) при $n \to \infty$ стремится к закону распределения χ^2 (см. лекцию 12) с числом степеней свободы $k = s - 1 - r$, где r – число параметров предполагаемого распределения, оцененных по данным выборки. Нормальное распределение характеризуется двумя параметрами, поэтому $k = s - 3$. Для выбранного критерия строится правосторонняя критическая область, определяемая условием

$$p(\chi^2 > \chi^2_{kp}(\alpha, k)) = \alpha,$$

где α – уровень значимости. Следовательно, критическая область задается неравенством $\chi^2 > \chi^2_{kp}(\alpha, k),$ а область принятия гипотезы - $\chi^2 < \chi^2_{kp}(\alpha, k)$.

Итак, для проверки нулевой гипотезы H_0: генеральная совокупность распределена нормально – нужно вычислить по выборке наблюдаемое значение критерия:

$$\chi^2_{набл} = \sum_{i=1}^{s} \frac{(n_i - n'_i)^2}{n'_i} \quad ,$$

а по таблице критических точек распределения χ^2 найти критическую точку $\chi^2_{кр}(\alpha, k)$, используя известные значения α и $k = s - 3$. Если $\chi^2_{набл} < \chi^2_{kp}$ - нулевую гипотезу принимают, при $\chi^2_{набл} > \chi^2_{kp}$ ее отвергают.

2. Проверка гипотезы о равномерном распределении.

При использовании критерия Пирсона для проверки гипотезы о равномерном распределении генеральной совокупности с предполагаемой плотностью вероятности

$$f(x) = \begin{cases} \dfrac{1}{b-a}, & x \in (a,b) \\ 0, & x \notin (a,b) \end{cases}$$

необходимо, вычислив по имеющейся выборке значение $\bar{x}_B$, оценить параметры *a* и *b* по формулам:

$$a^* = \bar{x}_B - \sqrt{3}\sigma_B, \;\; b^* = \bar{x}_B + \sqrt{3}\sigma_B \;\; ,$$

где *a** и *b** - оценки *a* и *b*. Действительно, для равномерного распределения $M(X) = \frac{a+b}{2}$, $\sigma(x) = \sqrt{D(X)} = \sqrt{\dfrac{(a-b)^2}{12}} = \dfrac{a-b}{2\sqrt{3}}$, откуда можно получить систему для определения *a** и *b**: $\begin{cases} \dfrac{b^* + a^*}{2} = \bar{x}_B \\ \dfrac{b^* - a^*}{2\sqrt{3}} = \sigma_B \end{cases}$, решением которой являются выражения .

Затем, предполагая, что $f(x) = \dfrac{1}{b^* - a^*}$, можно найти теоретические частоты по формулам

$$n'_1 = np_1 = nf(x)(x_1 - a^*) = n \cdot \frac{1}{b^* - a^*}(x_1 - a^*);$$

$$n'_2 = n'_3 = ... = n'_{s-1} = n \cdot \frac{1}{b^* - a^*}(x_i - x_{i-1}),\ \ i = 1,2,...,s-1;$$

$$n'_s = n \cdot \frac{1}{b^* - a^*}(b^* - x_{s-1}).$$

Здесь s – число интервалов, на которые разбита выборка.

Наблюдаемое значение критерия Пирсона вычисляется по формуле, а критическое – по таблице с учетом того, что число степеней свободы $k = s$-3. После этого границы критической области определяются так же, как и для проверки гипотезы о нормальном распределении.

3. Проверка гипотезы о показательном распределении.

В этом случае, разбив имеющуюся выборку на равные по длине интервалы, рассмотрим последовательность вариант $x_i^* = \frac{x_i + x_{i+1}}{2}$, равноотстоящих друг от друга (считаем, что все варианты, попавшие в i – й интервал, принимают значение, совпадающее с его серединой), и соответствующих им частот n_i (число вариант выборки, попавших в i – й интервал). Вычислим по этим данным $\bar{x}_B$ и примем в качестве оценки параметра λ величину $\lambda^* = \frac{1}{\bar{x}_B}$. Тогда теоретические частоты вычисляются по формуле

$$n'_i = n_i p_i = n_i p(x_i < X < x_{i+1}) = n_i (e^{-\lambda x_i} - e^{-\lambda x_{i+1}}).$$

Затем сравниваются наблюдаемое и критическое значение критерия Пирсона с учетом того, что число степеней свободы $k = s - 2$.

Критерий Колмогорова.

Этот критерий применяется для проверки простой гипотезы H_0 о том, что независимые одинаково распределенные случайные величины X_1, X_2, …, X_n имеют заданную непрерывную функцию распределения $F(x)$.

Найдем функцию эмпирического распределения $F_n(x)$ и будем искать границы двусторонней критической области, определяемой условием

$$D_n = \sup_{|x|<\infty} | F_n(x) - F(x) | > \lambda_n .$$

А.Н.Колмогоров доказал, что в случае справедливости гипотезы H_0 распределение статистики D_n не зависит от функции $F(x)$, и при $n \to \infty$

$$p(\sqrt{n} D_n < \lambda) \to K(\lambda), \;\; \lambda > 0,$$

где $K(\lambda) = \sum_{m=-\infty}^{\infty} (-1)^m e^{-2m^2\lambda^2}$ - критерий Колмогорова, значения которого можно найти в соответствующих таблицах. Критическое значение критерия $\lambda_n(\alpha)$ вычисляется по заданному уровню значимости α как корень уравнения $p(D_n \ge \lambda) = \alpha$.

Можно показать, что приближенное значение вычисляется по формуле

$$\lambda_n(\alpha) \approx \sqrt{\frac{z}{2n} - \frac{1}{6n}},$$

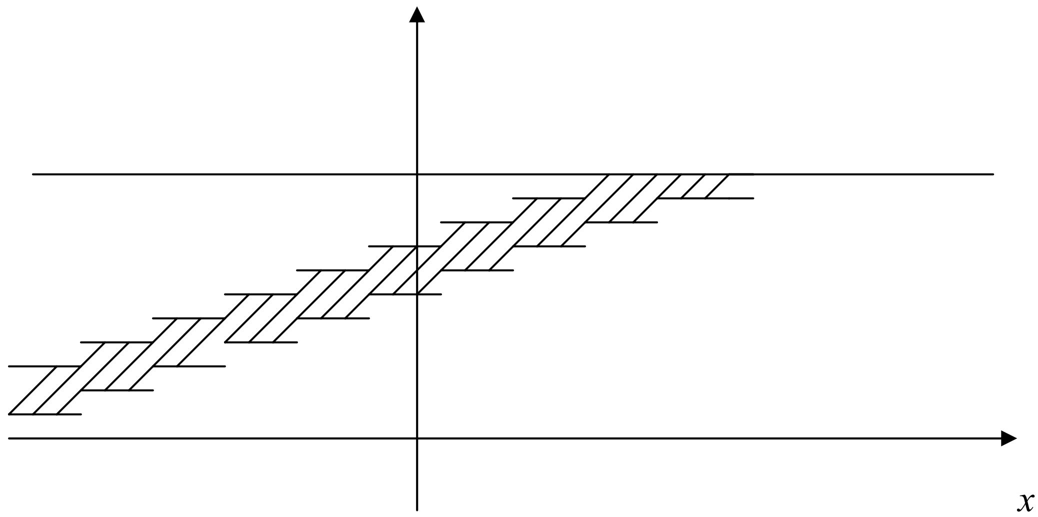
где z – корень уравнения $1 - K\left(\sqrt{\frac{\lambda}{2}}\right) = \alpha.$

На практике для вычисления значения статистики D_n используется то, что

$D_n = \max(D_n^+, D_n^-)$, где $D_n^+ = \max\limits_{1\le m \le n}\left(\frac{m}{n} - F(X_{(m)})\right), \;\; D_n^- = \max\limits_{1\le m \le n}\left(F(X_{(m)}) - \frac{m-1}{n}\right),$

а $X_{(1)} \le X_{(2)} \le ... \le X_{(n)}$ - вариационный ряд, построенный по выборке X_1, X_2, …, X_n.

Можно дать следующее геометрическое истолкование критерия Колмогорова: если изобразить на плоскости Оxy графики функций $F_n(x)$, $F_n(x) \pm\lambda_n(\alpha)$ (рис. 1), то гипотеза H_0 верна, если график функции $F(x)$ не выходит за пределы области, лежащей между графиками функций $F_n(x)$ -$\lambda_n(\alpha)$ и $F_n(x)$ +$\lambda_n(\alpha)$.



Приближенный метод проверки нормальности распределения, связанный с оценками коэффициентов асимметрии и эксцесса.

Определим по аналогии с соответствующими понятиями для теоретического распределения асимметрию и эксцесс эмпирического распределения.

<u>Определение.</u> Асимметрия эмпирического распределения определяется равенством

$$a_s = \frac{m_3}{\sigma_B^3},$$

где m_3 – центральный эмпирический момент третьего порядка.

Эксцесс эмпирического распределения определяется равенством

$$e_k = \frac{m_4}{\sigma_B^4} - 3,$$

где m_4 – центральный эмпирический момент четвертого порядка.

Как известно, для нормально распределенной случайной величины асимметрия и эксцесс равны 0. Поэтому, если соответствующие эмпирические величины достаточно малы, можно предположить, что генеральная совокупность распределения.

Содержание практических занятий

1. Вероятность случайного события.
2. Теоремы сложения и умножения вероятностей.
3. Формулы полной вероятности и Байеса.
4. Схема Бернулли.
5. Контрольная работа. Законы распределения дискретных случайных величин.
6. Законы распределения непрерывных случайных величин.
7. Числовые характеристики случайных величин.
8. Распределение Бернулли, Пуассона, геометрическое и другие распределения дискретных случайных величин.
9. Распределение непрерывных случайных величин.
10. Закон больших чисел.
11. Контрольная работа.
12. Метод наименьших квадратов.
13. Однофакторный анализ.
14. Точечные оценки параметров распределения.
15. Интервальные оценки параметров распределения.
16. Проверка статистических гипотез.
17. Проверка статистических гипотез.
18. Элементы теории корреляции.

Методические указание к практическим занятиям.

Теория вероятностей

Случайные события.

Цель изучения – ознакомиться с основными понятиями теории вероятностей и возможными областями применения аппарата теории вероятностей.

Данная тема включает в себя:

– основные понятия: событие; исход; достоверные, невозможные и случайные события; несовместные и совместные события; полная группа событий; противоположные события;

– классическое, геометрическое, статистическое определение вероятности события, их особенности, условия и возможные области применения;

– основные теоремы теории вероятностей (теоремы сложения для несовместных и совместных событий; теоремы умножения для независимых и зависимых событий), а также следствия из них (вероятность появления хотя бы одного события и формула полной вероятности). Формула Бейеса.

В результате изучения студент должен знать основные понятия и теоремы теории вероятностей, условия применения рассмотренных формул и теорем.

Изучив данную тему, студент должен уметь обосновывать способ решения и решать задачи, связанные с применением классического, геометрического и статистического определений вероятностей, а также теорем сложения и умножения вероятностей и следствий из них.

Вопросы для самопроверки.

1. Дайте определения достоверного, невозможного и случайного событий и приведите примеры этих событий.

2. Приведите примеры элементарных и сложных событий.

3. Сформулируйте классическое определение вероятности и условия его применения. Укажите интервал возможных значений вероятности.

4. В чем состоит отличие геометрического определения вероятности от классического определения?

5. Что называется относительной частотой появления события А?

6. Дайте определение несовместных и совместных событий, полной группы событий.

7. Какие события называются противоположными?

8. Что называется суммой событий?

9. В чем состоит отличие определения вероятности суммы событий для совместных и несовместных событий?

10. Что называется произведением событий?

11. Дайте определения независимых и зависимых событий, безусловной и условной вероятностей.

12. Приведите формулы для вычисления вероятности произведения независимых и зависимых событий, поясните отличие этих формул.

13. Как определить вероятность появления хотя бы одного события?

14. Сформулируйте формулу полной вероятности, поясните смысл входящих в нее величин.

15. Сформулируйте формулу Бейеса и поясните, в чем ее особенность.

Контрольные задачи по теме 1.

1. Чему равна вероятность того, что при бросании двух игральных косточек сумма выпавших очков равна 8 и их произведение равно 15?

Отв.: 1/18.

2. Из шести жетонов с буквами **и, к, л, м, н, о** случайным образом по одному, не возвращая обратно, извлекают пять букв, и выкладывают их в ряд. С какой вероятностью в результате этого опыта образуется слово "**лимон**"?

Отв.: 1/720.

3. В группе из 10 студентов 4 отличника. Из группы случайным образом отбираются 5 студентов. Найти вероятность того, что среди отобранных студентов окажутся 3 отличника?

Отв.: 5/21.

4. Три стрелка стреляют по мишени. Вероятность попадания в мишень для первого стрелка равна 0.9. для второго стрелка - 0,75 и для третьего 0,85. Найти вероятность того, что при стрельбе залпом в мишень попадут: а) все три стрелка; б) только два стрелка; в) не более одного стрелка.

Отв.: а) 0,57375; б) 0,35625; в) 0,07.

5. В группе спортсменов 25 футболистов, 10 хоккеистов и 15 бегунов. Вероятности выполнения квалификационного норматива соответственно равны для футболиста 0,9, для хоккеиста 0.8 и для бегуна 0.6. Какова вероятность того, что выбранный случайным образом спортсмен выполнит квалификационный норматив?

О*тв.: 0,79.*

6. Изделие проверяется на стандартность одним из двух товароведов. Вероятность того, что изделие попадет к первому товароведу, равна 0,55, а ко второму – 0,45. Вероятность того, что стандартное изделие будет признано стандартным первым товароведом, равна 0,9, а вторым – 0,98. Стандартное изделие при проверке было признано стандартным. Найти вероятность того, что это изделие проверил второй товаровед.

О*тв.: ≈ 0,47.*

Случайные величины

Цель изучения – ознакомиться с дискретными и непрерывными случайными величинами, способами их описания, законами распределения случайных величин и их числовыми характеристиками.

Данная тема включает в себя следующие вопросы:

– понятие дискретной случайной величины;

закон распределения дискретной случайной величины и способы его задания;

– формула Бернулли;

– биномиальный и пуассоновский законы распределения дискретных случайных величин;

– числовые характеристики (математическое ожидание, дисперсия, среднее квадратическое отклонение) дискретных случайных величин и методы их вычисления;

– понятие непрерывной случайной величины;

– функция распределения вероятностей случайной величины (интегральная функция распределения) и ее свойства;

– плотность распределения вероятностей непрерывной случайной величины (дифференциальная функция распределения) и ее свойства;

– связь между дифференциальной и интегральной функциями распределения непрерывной случайной величины;

– числовые характеристики непрерывных случайных величин;

– понятие моды и медианы непрерывной случайной величины;

– методы вычисления вероятности попадания значений непрерывной случайной величины в заданный интервал;

– равномерное распределение вероятностей непрерывной случайной величины (дифференциальная и интегральная функции распределения; числовые характеристики; определение вероятности попадания случайной величины в заданный интервал);

– нормальное распределение вероятностей непрерывной случайной величины (дифференциальная и интегральная функции распределения; нормированное нормальное распределение; функция Лапласа; числовые характеристики; определение вероятности попадания случайной величины в заданный интервал; правило трех сигм);

– экспоненциальное (показательное) распределение вероятностей непрерывной случайной величины (дифференциальная и интегральная функции распределения; числовые характеристики; определение вероятности попадания случайной величины в заданный интервал).

Вопросы для самопроверки.

1. Приведите примеры дискретных случайных величин.
2. Что называется законом распределения дискретной случайной величины и как его можно задавать?
3. Укажите условия применения формулы Бернулли.
4. Назовите общие и отличительные черты биномиального и пуассоновского распределений дискретных случайных величин.
5. Дайте определения математического ожидания, дисперсии, среднего квадратического отклонения дискретных случайных величин и

приведите формулы для их вычисления. Поясните сущность числовых характеристик случайных величин.

6. В чем отличие непрерывных случайных величин от дискретных случайных величин? Приведите примеры непрерывных случайных величин.

7. Как задается закон распределения непрерывной случайной величины?

8. В чем заключается сущность интегральной и дифференциальной функций распределения непрерывных случайных величин, и как они связаны между собой?

9. Приведите формулы для нахождения числовых характеристик непрерывных случайных величин.

10. Дайте характеристику равномерного закона распределения непрерывных случайных величин.

11. Приведите формулы для нахождения числовых характеристик равномерно распределенных непрерывных случайных величин.

12. Дайте характеристику нормального закона распределения непрерывных случайных величин.

13. Назовите параметры нормального закона распределения и поясните их смысл.

14. В чем состоит отличие нормированного нормального распределения от нормального распределения? Приведите формулы для дифференциальной и интегральной функций распределения нормального и нормированного нормального законов распределения.

15. Что называется функцией Лапласа? Укажите ее свойства и приведите примеры ее использования.

16. Сформулируйте правило трех сигм.

17. Дайте характеристику экспоненциального (показательного) закона распределения непрерывных случайных величин. Что является параметром показательного распределения?

18. Приведите формулы для нахождения числовых характеристик непрерывных случайных величин, распределенных по показательному закону.

Контрольные задачи по теме 2.

1. Случайная величина X задана на всей числовой оси Ox функцией распределения $F(x) = 1/2 + (\mathrm{arctg}\ x)/\pi$. Найти вероятность того, что в результате испытания величина X примет значение, заключенное в интервале (0, 1). *Отв.: 0,25.*

2. Случайная величина X задана интегральной функцией распределения

$$F(x) = \begin{cases} 0 & \text{при } x \le 0, \\ x^2 & \text{при } 0 < x \le 1, \\ 1 & \text{при } x > 1. \end{cases}$$

Найти дифференциальную функцию распределения и числовые характеристики случайной величины X.

Отв.:

$$f(x) = \begin{cases} 0 & \text{при } x \le 0, \\ 2x & \text{при } 0 < x \le 1, \\ 0 & \text{при } x > 1. \end{cases}$$

$$M(X) = \frac{2}{3}; \quad D(X) = \frac{1}{18}; \quad \sigma(X) = \frac{1}{3\sqrt{2}}.$$

3. Непрерывная случайная величина X задана дифференциальной функцией распределения $f(x) = \sin x$ в интервале $(0, \pi/2)$; вне этого интервала $f(x) = 0$. Найти интегральную функцию распределения и математическое ожидание случайной величины X, а также вероятность того, что в результате испытания величина X примет значение, принадлежащее интервалу $(\pi/3, \pi/2)$.

Отв.:

$$F(x) = \begin{cases} 0 & \text{при } x \le 0, \\ 1 - \cos x & \text{при } 0 < x \le \pi/2, \\ 1 & \text{при } x > \pi/2. \end{cases}$$

$M(X) = 1;\ P(\pi/3 < x < \pi/2) = 0{,}5.$

4. Найти дифференциальную и интегральную функции распределения, а также математическое ожидание случайной величины X, равномерно распределенной в интервале (20, 30).

Отв.:

$$f(x) = \begin{cases} 0 & \text{при } x \le 20, \\ 0{,}1 & \text{при } 20 < x \le 30, \\ 0 & \text{при } x > 30. \end{cases}$$

$$F(x) = \begin{cases} 0 & \text{при } x \le 20, \\ \dfrac{x-20}{10} & \text{при } 20 < x \le 30, \\ 1 & \text{при } x > 30. \end{cases}$$

M(X) = 25.

5. Докажите, что для вычисления дисперсии случайной величины, равномерно распределенной в интервале (a, b), может быть использована формула $D(x) = (b – a)^2/12$.

6. Найти дисперсию и среднее квадратическое отклонение случайной величины X, равномерно распределенной в интервале (100, 106).

Отв.: 3: $\sqrt{3}$.

7. Найти дифференциальную функцию распределения и числовые характеристики случайной величины X, распределенной по нормальному закону с параметрами a = – 10 и σ=3.

Отв.:

$$f(x) = \frac{1}{3\sqrt{2\pi}} e^{-\frac{(x+10)^2}{18}}$$

M(X) = – 10; D(X) = 9; σ(X) *=3.*

8. Математическое ожидание и среднее квадратическое отклонение нормально распределенной случайной величины X соответственно равны 30 и 10. Найти вероятность того, что в результате испытания величина X примет значение, принадлежащее интервалу (25, 40).

Отв.: 0,5328.

9. Производятся измерения диаметра вала без систематических ошибок. Случайные ошибки измерения X подчинены нормальному закону распределения со средним квадратическим отклонением σ=10 мм. Найти вероятность того, что измерение будет произведено с ошибкой, не превосходящей по абсолютной величине 15 мм.

Отв.: 0,8664.

10. Случайная величина X распределена нормально с математическим ожиданием a = 10 и средним квадратическим отклонением σ = 5. Найти интервал, симметричный относительно математического ожидания, в который с вероятностью 0,9973 попадет величина X в результате испытания. *Отв.: (–5; 25).*

11. Написать дифференциальную и интегральную функции распределения для непрерывной случайной величины, распределенной по показательному закону с параметром $\lambda = 0{,}1$.

Отв.: $$f(x) = \begin{cases} 0 & \text{ïðè } x < 0, \\ 0{,}1e^{-0{,}1x} & \text{ïðè } x \geq 0. \end{cases}$$

$$F(x) = \begin{cases} 0 & \text{ïðè } x < 0, \\ 1 - e^{-0{,}1x} & \text{ïðè } x \geq 0. \end{cases}$$

12. Дифференциальная функция распределения непрерывной случайной величины X, распределенной по показательному закону, $f(x) = 0{,}01 \bullet e^{-0{,}01x}$ при $x \geq 0$ и $f(x) = 0$ при $x<0$. Найти вероятность того, что в результате испытания X попадет в интервал (1, 3). *Отв.:* $\approx 0{,}02$.

Основы математической статистики

Цель изучения – ознакомиться с основными понятиями математической статистики и возможными областями применения методов математической статистики.

Данная тема включает в себя следующие вопросы:

- цели и задачи математической статистик**и;**

- выборочный метод: сущность и основные понятия;

- статистическое распределение выборки и формы его представления;

- эмпирическая функция распределения;

- полигон и гистограмма;

- статистическое оценивание параметров распределения (точечные и интервальные оценки);

- числовые характеристики выборки (средняя выборочная; выборочная и исправленная выборочная дисперсии; выборочное и исправленное выборочное среднее квадратическое отклонение);

- метод произведений и метод сумм нахождения числовых характеристик выборки.

Вопросы для самопроверки.

1. Назовите основные задачи математической статистики.
2. Дайте определения понятий: генеральная и выборочная совокупности; объем генеральной и выборочной совокупности.
3. Как обеспечивается репрезентативность выборки? Назовите основные методы отбора объектов в выборочную совокупность.
4. Укажите основные виды представления результатов выборочных наблюдений.
5. Как найти эмпирическую функцию распределения?
6. В чем отличие полигона и гистограммы распределения?
7. Приведите формулы для вычисления числовых характеристик выборки и укажите особенности их применения.
8. Назовите основные свойства точечных оценок.
9. Сравнительный анализ преимуществ и недостатков точечных и интервальных оценок.
10. Приведите формулу нахождения доверительного интервала для математического ожидания нормально распределенной случайной величины по результатам выборочных наблюдений при известном среднем квадратическом отклонении генеральной совокупности.
11. Приведите примеры использования метода сумм и метода произведений для определения числовых характеристик выборки.

Контрольные задачи по теме 3.

1. Выборка задана в виде распределения частот:

x_i	–1	0	3	7	10
n_i	4	7	11	2	1

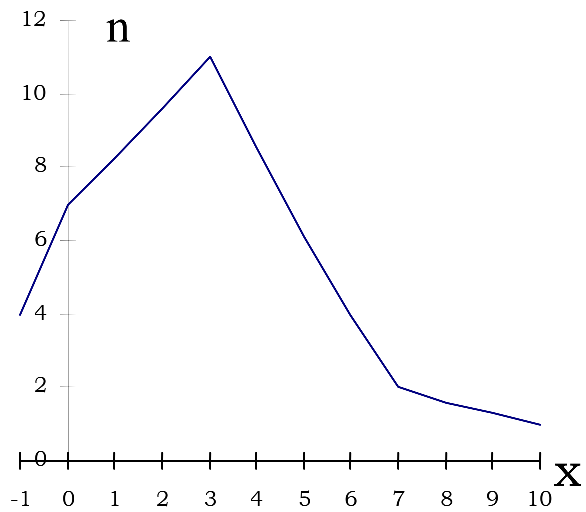
Найти: а) распределение относительных частот, б) эмпирическую функцию распределения. Построить: в) полигон частот и г) полигон относительных частот.

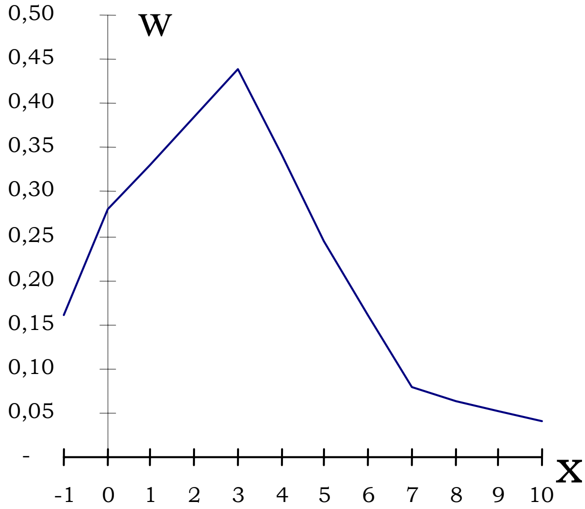
Отв.: а)

x_i	-1	0	3	7	10
w_i	$0{,}16$	$0{,}28$	$0{,}44$	$0{,}08$	$0{,}04$

б)

$$F^*(x) = \begin{cases} 0 & \text{при } x \le -1 \\ 0{,}16 & \text{при } -1 < x \le 0 \\ 0{,}44 & \text{при } 0 < x \le 3 \\ 0{,}88 & \text{при } 3 < x \le 7 \\ 0{,}96 & \text{при } 7 < x \le 10 \\ 1{,}00 & \text{при } x > 10 \end{cases}$$

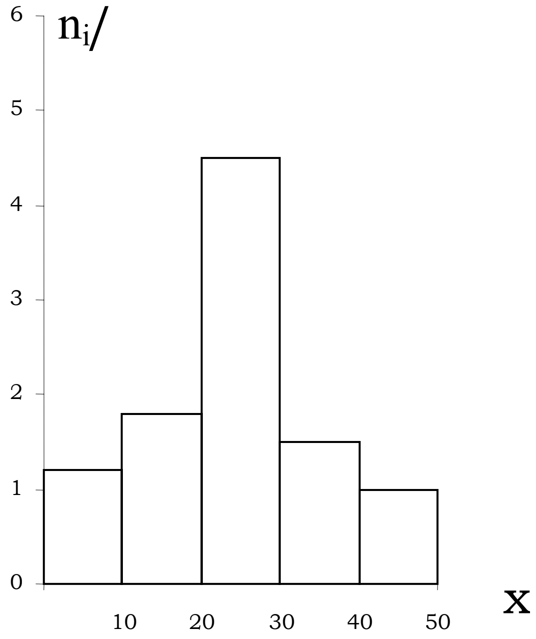


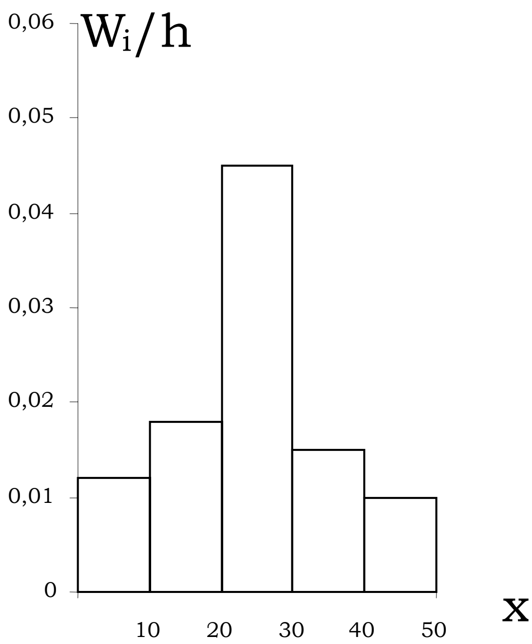


2. Результаты выборочных наблюдений над непрерывной случайной величиной X приведены ниже в виде интервалов одинаковой длины и соответствующих им частот:

$x_i - x_{i+1}$	0 – 10	10 – 20	20 – 30	30 – 40	40 – 50
n_i	12	18	45	15	10

Построить: а) гистограмму частот и б) гистограмму относительных частот случайной величины X.





3. Найти выборочную среднюю, выборочную дисперсию и выборочное среднее квадратическое отклонение случайной величины X, если распределение выборки имеет следующий вид:

x_i	–5	1	3	6	15
n_i	8	14	12	9	7

Отв.: $\bar{x}_в = 3,38$; $D_в \approx 33$; $\sigma_в \approx 5,74$.

4. Найти несмещенные оценки генеральной средней и генеральной дисперсии, если по результатам выборочных наблюдений получено следующее распределение

x_i - x_{i+1}	0 – 8	8 – 16	16 – 24	24 – 32	32 – 40
n_i	5	10	20	8	7

Отв.: $\bar{x}_в = 20,32$; $S^2(X) \approx 86,10$.

5. Найти интервальную оценку с доверительной вероятностью 0,95 для неизвестного математического ожидания нормально распределенной случайной величины X, если известны: генеральное среднее квадратическое отклонение

$\sigma(X) = 4$; выборочная средняя $\bar{x}_Â$ = 30; объем выборки n = 25.

Отв.: $28,432 < a < 31,568$.

6. С помощью метода произведений найти выборочную среднюю, выборочную дисперсию и выборочное среднее квадратическое отклонение случайной величины X, если распределение выборки имеет следующий вид:

x_i - x_{i+1}	–10 – 0	0 – 10	10 – 20	20 – 30	30 – 40	40 – 50	50 – 60
n_i	7	12	19	24	17	11	10

Отв.: $\bar{x}_Â = 25,5$; $D_Â = 280,75$; $\sigma_в \approx 16,76$.

Статистическая проверка статистических гипотез

Цель изучения – ознакомиться с основными понятиями и методами статистической проверки статистических гипотез.

Данная тема включает в себя следующие вопросы:

– понятие статистической гипотезы;

– виды статистических гипотез;

– общий порядок статистической проверки статистических гипотез;

– статистические критерии;

– ошибки при статистической проверке статистических гипотез;

– проверка статистических гипотез о неизвестных параметрах известных законов распределения генеральных совокупностей

– сравнение двух средних генеральных совокупностей;

– сравнение двух дисперсий нормальных генеральных совокупностей;

– сравнение дисперсий нескольких нормальных генеральных совокупностей;

– проверка статистических гипотез о неизвестных законах распределения генеральных совокупностей:

– проверка гипотезы о показательном распределении генеральной совокупности;

– проверка гипотезы о нормальном распределении генеральной совокупности;

– проверка гипотезы о равномерном распределении генеральной совокупности;

– проверка гипотезы о биномиальном распределении генеральной совокупности.

Вопросы для самопроверки.

1. Дайте определения статистической гипотезы, нулевой и альтернативной гипотез.
2. Почему проверка статистических гипотез называется статистической?
3. Назовите виды статистических гипотез и приведите примеры таких гипотез.
4. Какие исходы возможны при проверке статистических гипотез? Дайте определения ошибкам первого и второго родов.
5. Что называется статистическим критерием?
6. Сформулируйте понятия области принятия гипотезы и критической области. От чего зависит выбор типа критической области (левосторонней, правосторонней, двухсторонней), и как это связано с видом альтернативной гипотезы?
7. В чем заключается порядок статистической проверки статистических гипотез?
8. Что называется уровнем значимости при проверке статистических гипотез, и с какой вероятностью принимается решение?
9. Какой критерий применяется при проверке гипотезы о равенстве генеральных дисперсий двух нормальных генеральных совокупностей? Как вычисляется наблюдаемое значение этого критерия?
10. Приведите формулу для определения наблюдаемого значения критерия при сравнении средних двух нормальных генеральных совокупностей при известных генеральных дисперсиях и в случае больших независимых выборок.
11. Как проверить гипотезу о равенстве дисперсий нескольких нормально распределенных генеральных совокупностей в случае по выборкам одинакового объема? Что следует использовать в качестве оценки генеральной дисперсии, если выдвинутая гипотеза будет подтверждена?

12. В чем состоит сущность критерия Пирсона при проверке гипотезы о виде закона распределения генеральной совокупности? Приведите формулу для определения наблюдаемого значения критерия Пирсона.

13. В чем состоит отличие проверки гипотезы о виде закона распределения в случае различных законов распределения? Поясните это на примерах.

Контрольные задачи по теме 4.

1. По двум независимым выборкам, объемы которых $n_1 = 9$ и $n_2 = 16$, извлеченным из нормальных генеральных совокупностей X и Y, найдены исправленные выборочные дисперсии $S^2(X) = 34{,}02$ и $S^2(Y) = 12{,}25$. Проверить нулевую гипотезу H_0: D(X) = D(Y) о равенстве генеральных дисперсий при конкурирующей гипотезе H_1: D(X) > D(Y) и уровне значимости $\alpha = 0{,}01$.

Отв.: Так как $F_{набл}$ = 2,8 меньше $F_{кр}$(0,01; 8; 15) = 4, то нет оснований отвергнуть нулевую гипотезу.

2. По двум независимым выборкам, объемы которых $n_1 = 14$ и $n_2 = 10$, извлеченным из нормальных генеральных совокупностей X и Y, найдены исправленные выборочные дисперсии $S^2(X) = 0{,}84$ и $S^2(Y) = 2{,}52$. Проверить нулевую гипотезу H_0: D(X) = D(Y) о равенстве генеральных дисперсий при конкурирующей гипотезе H_1: D(X) $\neq$ D(Y) и уровне значимости $\alpha = 0{,}1$. *Отв.: Так как $F_{набл}$ =3 больше $F_{кр}$(0,05; 9; 13) = 2,72, то нулевую гипотезу отвергаем.*

3. Четыре фасовочных автомата настроены на отвешивание одного и того же веса. На каждом автомате отвесили по 10 проб, а затем эти же пробы

взвесили на точных весах и нашли по полученным отклонениям исправленные выборочные дисперсии: 0,012; 0,021; 0,025; 0,032. Можно ли при уровне значимости α = 0,05 считать, что автоматы обеспечивают одинаковую точность взвешивания (т.е. проверить нулевую гипотезу о равенстве генеральных дисперсий), и, в случае подтверждения данной гипотезы, найти оценку генеральной дисперсии.

Отв.: а) так как $G_{набл}$ *= 0,3556 меньше* $G_{кр}(0,05;\ 9;\ 4) = 0,5017$*, то можно считать, что автоматы обеспечивают одинаковую точность взвешивания; б)* $S^2(X) = 0,0225$.

4. По четырем независимым выборкам, объемы которых $n_1 = 17$, $n_2 = 20$, $n_3 = 15$, $n_4 = 16$, извлеченным из нормальных генеральных совокупностей, найдены исправленные выборочные дисперсии, соответственно равные 2,5; 3,6; 4,1; 5,8. Требуется: а) при уровне значимости α = 0,05 проверить нулевую гипотезу об однородности дисперсий; б) оценить генеральную дисперсию.

Отв.: а) $k = 64$; $\sum k_i * S_i^2 = 252,8$; $\sum k_i * \lg S_i^2 = 36,9663$; $V = 2,8$; $B_{набл} < 2,8$; $\chi^2_{кр}(0,05;\ 3) = 7,8$; $B_{набл} < \chi^2_{кр}(0,05;\ 3)$. *Следовательно, нет оснований отвергнуть гипотезу об однородности дисперсий; б)* $S^2(X) = 4,00$.

5. По выборке объемом n = 30 найден средний вес $\bar{x}_в = 130$ деталей, изготовленных на первом станке; по выборке объемом m = 40 найден средний вес $\bar{y}_в = 125$ деталей, изготовленных на втором станке. Известны генеральные дисперсии D(X) = 60 $г^2$, D(Y) = 80 $г^2$. Требуется при уровне значимости α = 0,05 проверить нулевую гипотезу H_0: M(X) = M(Y) о равенстве генеральных средних данных совокупностей при альтернативной

гипотезе H_1: $M(X) \neq M(Y)$. Предполагается, что случайные величины X и Y распределены нормально и выборки независимы.

Отв.: Так как $Z_{набл} = 2,5$ больше $Z_{кр} = 1,96$, то нулевую гипотезу отвергаем (т.е. средний вес деталей различается значимо).

6. По выборке объемом n = 50 найден средний размер $\bar{x}_в = 20,1$ мм диаметра валиков, изготовленных автоматом №1; по выборке объемом m = 50 найден средний размер $\bar{y}_в = 19.8$ мм диаметра валиков, изготовленных автоматом №2. Генеральные дисперсии известны: D(X) = 1,750 мм^2 и D(Y) = 1,375 мм^2. Требуется при уровне значимости $\alpha = 0,05$ проверить нулевую гипотезу H_0: M(X) = M(Y) о равенстве средних размеров валиков, изготовленных на данных автоматах, при альтернативной гипотезе H_1: M(X) > M(Y). Предполагается, что случайные величины X и Y распределены нормально и выборки независимы.

Отв.: Так как $Z_{набл} = 1,2$ меньше $Z_{кр} = 1,645$, то нет оснований отвергать нулевую гипотезу.

7. Используя критерий Пирсона, при уровне значимости 0,01, установить, случайно или значимо расхождение между эмпирическими частотами n_i и теоретическими частотами n_i', вычисленными исходя из гипотезы о нормальном распределении генеральной совокупности X:

n_i	8	16	40	72	36	18	10
n_i'	6	18	36	76	39	18	7

Отв.: Так как $\chi^2_{набл} = 3,061$ меньше $\chi^2_{кр}(0,01;4) = 13,3$, то расхождение между эмпирическими и теоретическими частотами является случайным и, *следовательно, гипотезу о нормальном распределении генеральной совокупности X нет оснований отвергать.*

8. Используя критерий Пирсона, при уровне значимости 0,05, проверить гипотезу о нормальном распределении генеральной совокупности X, если выборочное распределение из этой совокупности имеет следующий вид;

x_i	5	7	9	11	13	15	17	19	21
n_i	15	26	25	30	26	21	24	20	13

Указание: при определении выборочной средней и выборочного среднего квадратического отклонения использовать метод произведений.

Отв.: Так как $|\chi^2_{набл}| = 22{,}2$ *больше* $\chi^2_{кр}(0{,}05;\ 6) = 12{,}6$, *то нулевую гипотезу о нормальном распределении генеральной совокупности X отвергаем.*

9. Используя критерий Пирсона, при уровне значимости 0,05, проверить гипотезу о нормальном распределении генеральной совокупности X, если выборочное распределение из этой совокупности имеет следующий вид;

x_i - x_{i+1}	(–20) – (–10)	(–10) – 0	0 – 10	10 – 20	20 – 30	30 – 40	40 – 50
n_i	20	47	80	89	40	16	8

Отв.: Так как $\chi^2_{набл} = 29{,}664$ *больше* $\chi^2_{кр}$ $(0{,}05;\ 4) = 9{,}5$, *то нулевую гипотезу о нормальном распределении генеральной совокупности X отвергаем.*

10. В итоге испытаний 1000 элементов на время безотказной работы получено эмпирическое распределение, приведенное в таблице (в верхней графе указаны интервалы времени в часах; в нижней графе – частоты n_i, т. е. количество элементов, отказавших в *i*–м интервале).

x_i - x_{i+1}	0– 10	10 – 20	20 – 30	30 – 40	40 – 50	50 – 60	60 – 70
n_i	365	245	150	100	70	45	25

Требуется при уровне значимости α = 0,01 проверить нулевую гипотезу о том, что время безотказной работы элементов подчиняется показательному закону распределения.

Отв.: Так как $\chi^2_{набл}$ *=15,88 больше* $\chi^2_{кр}$*(0,01; 5) = 15,1, то нулевую гипотезу о показательном распределении времени безотказной работы X отвергаем*

11. В некоторой местности в течение 300 суток регистрировалась среднесуточная температура воздуха. В итоге наблюдений было получено эмпирическое распределение, приведенное ниже в таблице (в первом столбце указан интервал температуры в градусах, во втором столбце – частота n_i, т.е. количество дней, среднесуточная температура воздуха которых принадлежит этому интервалу).

$x_i - x_{i+1}$	n_i		$x - x_{i+1}$	n_i
–40 – (–30)	25		0 – 10	40
–30 – (–20)	40		10 – 20	46
–20 – (–10)	30		20 – 30	48
–10 – 0	45		30 - 40	26

Требуется при уровне значимости α = 0,05 проверить гипотезу о том, что среднесуточная температура воздуха распределена равномерно.

Отв.: Так как $\chi^2_{набл}$ *= 7,71 меньше* $\chi^2_{кр}$ *(0,05; 5) = 11,1, то нет оснований отвергать нулевую гипотезу о равномерном распределении среднесуточной температуры воздуха X.*

12. Опыт, состоящий в одновременном подбрасывании четырех монет, повторили 100 раз, эмпирическое распределение дискретной случайной

величины X – числа появившихся “гербов” – оказалось следующим (в первой строке указано число x_i выпавших “гербов” в одном бросании монет; во второй строке – частота n_i, т.е. число бросаний, при которых выпало x_i “гербов”):

x_i	0	1	2	3	4
n_i	8	20	42	22	8

Требуется при уровне значимости α = 0,05 проверить гипотезу о том, что случайная величина X распределена по биномиальному закону.

Указание. Принять вероятность выпадения “герба” при бросании одной монеты p = 0,5.

Отв.: Так как $\chi^2_{набл}$ *= 2,88 меньше* $\chi^2_{кр}$*(0,05; 4) = 9,5, то нет оснований отвергать нулевую гипотезу о биномиальном распределении случайной величины X.*

13. Отдел технического контроля проверил n = 100 партий изделий по N = 10 изделий в каждой партии и получил следующее эмпирическое распределение дискретной случайной величины X – числа нестандартных изделий (в первой строке указано число x_i нестандартных изделий в одной партии; во второй строке – частота n_i, т.е. количество партий, содержащих x_i нестандартных изделий):

x_i	0	1	2	3	4	5	6	7
n_i	2	3	10	22	26	20	12	5

Требуется при уровне значимости α = 0,01 проверить гипотезу о том, что случайная величина X распределена по биномиальному закону.

Указания.

1. Найти сначала относительную частоту появления нестандартных изделий (отношение числа всех нестандартных изделий к общему числу проверенных изделий) и принять ее в качестве оценки p^*

вероятности того, что наудачу взятое изделие окажется нестандартным.

2. При составлении расчетной таблицы для сравнения эмпирических и теоретических частот с помощью критерия Пирсона необходимо для первых двух групп объединить эмпирические частоты (2 + 3 = 5) , следовательно, число групп после объединения составит s = 7.
3. Один параметр (вероятность p) оценивался по выборке, поэтому при определении числа степеней свободы надо вычесть из s не единицу, а два: k = s – 2 = 5.

Отв.: $p^* = 0{,}4$; *так как* $\chi^2_{набл} = 0{,}68$ *меньше* $\chi^2_{кр}(0{,}01;\ 5) = 15{,}1$, *то нет оснований отвергать нулевую гипотезу о биномиальном распределении случайной величины* X.

14. В итоге проверки на нестандартность 200 ящиков консервов получено следующее эмпирическое распределение (в первой строке указано число x_i нестандартных коробок консервов в одном ящике; во второй строке – частота n_i, т.е. число ящиков, содержащих x_i нестандартных консервов):

x_i	0	1	2	3	4
n_i	132	43	20	3	2

Требуется при уровне значимости $\alpha = 0{,}05$ проверить гипотезу о том, что случайная величина X – число нестандартных коробок консервов распределена по закону Пуассона.

Указание. Объединить малочисленные частоты двух последних групп.

Отв.: Так как $\chi^2_{набл} = 9{,}27$ больше $\chi^2_{кр}(0{,}05;\ 2) = 6{,}0$, то нулевая гипотеза о распределении случайной величины X по закону Пуассона отвергается.

График самостоятельной работы

№ п/п	Содержание сам. работы	Объем в час.	Номер недели																		Форма контроля
			1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	
1	Подготовка к лекциям	17	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	экзамен
2	Выполнение дом. зада ний	25	0	1,5	1,5	1,5	1,5	1,5	1,5	1,5	1,5	1,5	1,5	1,5	1,5	1,5	1,5	1,5	1,5	1	Проверка дом. зад.
3	Выполнение РГР № 1	10	1	1	1	1	1	1	1	1	1	1									защита
4	Выполнение РГР № 2	7											1	1	1	1	1	1	1		защита
5	Подготовка к контрольной работе				2								2								Проверка кон.

Методические указания для выполнения домашних и контрольных работ

Тема 1. Материал данной темы является вспомогательным при изучении теории вероятностей. Правила и формулы комбинаторики используются при непосредственном вычислении вероятностей. Поэтому, в первую очередь, необходимо изучить и усвоить основные виды соединений: размещения, перестановки и сочетания. Надо знать формулы для их вычисления и уметь их применять на практике.

Тема 2. При изучении темы 2 необходимо, прежде всего, усвоить первоначальные понятия и определения, на которые опирается теория вероятностей. К таким понятиям относятся понятия события и равно возможности.

Необходимо рассмотреть различные виды событий и операции над событиями, обратите внимание на понятия совместных и несовместных событий, противоположных событий.

Вероятность события есть численная мера объективной возможности его появления. Следует хорошо изучить классическое определение вероятности, понятие статистической и геометрической вероятностей, условия применения того или иного определения вероятностей.

Тема 3. Теория вероятностей позволяет определить вероятность события по известным вероятностям других событий, если они связаны с данным событием. В этом случае используют теоремы сложения и умножения вероятностей. Они дают возможность найти вероятность появления одного из нескольких случайных событий или вероятность совместного наступления двух и более событий.

При изучении теоремы сложения вероятностей следует обратить внимание на случаи совместных и несовместных событий, т.к. теорема для этих случаев формулируется и применяется по-разному.

Надо иметь в виду, что есть отличия в формулировке и применении теоремы умножения вероятностей для зависимых и независимых событий. Поэтому необходимо внимательно изучить определения зависимых и независимых событий, а также понятие условной вероятности.

Формула полной вероятности является следствием теоремы сложения и умножения вероятностей и позволяет определить вероятность события, которое может произойти с одним из несовместных событий, образующих полную группу, вероятности которых известны.

Формулы Байеса позволяют “пересмотреть” вероятности гипотез после того, как становятся известными итоги опыта, в результате которого появилось данное событие.

Формулы Байеса находят все большее применение при решении проблем управления в экономике, связанных с недостаточной информацией.

Тема 4. Часто встречаются задачи, в которых требуется определить вероятность **К** появлений события **А** в результате **n** испытаний. Данная тема посвящена рассмотрению таких задач для случая независимых испытаний, которые могут повторяться многократно.

Следует иметь в виду, что формула Бернулли обычно применяется, когда число испытаний невелико (менее 10). При большом числе испытаний ее применение приводит к громоздким вычислениям. В таких случаях применяют асимптотические формулы. Одну из них дает локальная теорема Муавра-Лапласа.

С увеличением числа испытаний точность значений вероятности возрастает. Однако, если вероятность наступления события в одном испытании близка к нулю, то значения вероятностей, вычисляемых по формуле Муавра-Лапласа, оказывается весьма приближенным. В этом случае применяется формула Пуассона.

Если требуется найти вероятность того, что число наступлений события в **n** независимых испытаниях заключено в некоторых границах, то применяется интегральная теорема Муавра-Лапласа.

В результате изучения материала данной темы необходимо знать основные формулы, условия их применения и уметь применять их при решении практических задач.

Краткие методические указания по изучению материала

Тема 1. Понятие случайной величины – одно из основных в теории вероятностей. Случайная величина является числовой характеристикой результата эксперимента. Поэтому материал данной темы является важнейшим с точки зрения освоения теории вероятностей и математической статистики.

При изучении темы необходимо разобрать и усвоить понятия дискретных и непрерывных случайных величин, законы их распределения.

Подробно изучить интегральную и дифференциальную функции распределения, их свойства и применение.

Следует иметь в виду, что удобной формой отражения наиболее существенных особенностей случайной величины являются числовые характеристики. К ним относятся математическое ожидание, дисперсия, моменты различных порядков.

Необходимо изучить и усвоить определения основных числовых характеристик и их свойства для дискретных и непрерывных случайных величин.

Вместе с тем надо хорошо знать некоторые виды распределений:

биномиальное, геометрическое, распределение Пуассона, равномерное, показательное и нормальное.

Тема 2. Материал темы изучается обзорно в ознакомительном порядке. При изучении результатов наблюдений над реальными массовыми случайными явлениями имеют место некоторые закономерности. Следует обратить внимание на то, что они обладают свойством устойчивости. Суть этого свойства состоит в том, что конкретные особенности каждого отдельного случайного явления почти не сказываются на среднем результате большой массы подобных явлений.

Характеристики случайных событий и случайных величин, наблюдаемых в испытаниях, при неограниченном увеличении числа испытаний становятся практически не случайными.

Предельные теоремы теории вероятностей устанавливают зависимость между случайностью и необходимостью. По смыслу их можно разбить на две группы, одна из которых называется законом больших чисел, а другая – центральной предельной теоремой.

Закон больших чисел – это обобщенное название нескольких теорем, из которых следует, что при неограниченном увеличении числа испытаний средние величины стремятся к некоторым постоянным, из этих теорем необходимо знать формулировку и практическое применение теоремы Чебышева и теоремы Бернулли.

Центральная предельная теорема устанавливает связь между законом распределения суммы случайных величин и его предельной формой – нормальным законом распределения.

Тема 3. Прежде всего, необходимо рассмотреть предмет и задачи математической статистики. Надо иметь в виду, что основой математической статистики является выборочный метод. Поэтому следует обратить внимание на характеристики и свойства выборки, способы осуществления выборки из генеральной совокупности.

Следующая задача математической статистики заключается в обработке результатов наблюдений. Необходимо рассмотреть операции ранжирования и группировки опытных данных, дискретный и интервальный вариационные ряды и способы их построения.

При рассмотрении аналогов интегральной и дифференциальной функций распределения, а также выборочных средних и выборочных дисперсий следует опираться на соответствующие понятия теории вероятностей.

Тема 4. При изучении данной темы следует иметь в виду, что любая выборка несет в себе ограниченную информацию о поведении генеральной совокупности. Поэтому в качестве числовых характеристик случайной величины принимают приближенные значения, определяемые по выборке.

Такие значения называют точечными оценками числовых характеристик. Следует внимательно изучить свойства точечных оценок и методы их получения (метод моментов, метод максимального правдоподобия). Рассмотреть эти методы на примерах, представленных в [1] и [2].

Студент должен иметь представление об интервальных оценках и уметь строить доверительные интервалы для оценки параметров нормального распределения.

Тема 5. В данной теме рассматриваются два вида статистических гипотез: о предполагаемых значениях параметров известного распределения и о виде предполагаемого распределения.

При этом студент должен знать понятия основной и альтернативной гипотезы, определения ошибок первого и второго рода, понятие критической области, а также основные этапы проверки гипотезы. В качестве основных примеров следует рассмотреть проверку гипотез о числовых значениях параметров нормального распределения (математического ожидания и дисперсии) и проверку гипотезы о модели закона распределения с применением критерия

Тема 6. При изучении данной темы, прежде всего, необходимо уяснить понятия функциональной, стохастической и корреляционной зависимостей между случайными величинами. Следует иметь в виду, что в качестве количественной меры зависимости случайных величин используют корреляционный момент или коэффициент корреляции, а в качестве математической модели этой зависимости применяют уравнение регрессии.

Студент должен уметь по данным корреляционной таблицы находить выборочный коэффициент корреляции и строить уравнения прямой линии регрессии.

Задания для выполнения домашних и

расчетно - графических работ включены в пособиях

.

1. Торопчина Г. Н., Вохминцева Г. П., Шевченко И. Н. Введение в теорию вероятностей. - Благовещенск.: АмГУ, 2006.
2. Торопчина Г. Н., Вохминцева Г. П., Шевченко И. Н. Лабораторные работы по математической статистике. - Благовещенск.: АмГУ, 2006.

Примерные задания для контрольных и расчетно-графических работ

Самостоятельная работа

1. Вычислите A_6^2, C_8^5, P_7

2. Дано: $P(A \cap B) = 0{,}4$; $P_B(A) = \dfrac{2}{3}$; $P_A(B) = \dfrac{3}{4}$. Найдите $P(A)$, $P(B)$, $P(A \cup B)$ и выясните, зависимы ли события *A, B*.

3. Покупатель может приобрести акции двух компаний – *A* и *B*. Надежность первой оценивается экспертами на уровне 90% , второй – 80%. Чему равна вероятность того, что: а) обе компании в течение года не станут банкротами; б) наступит хотя бы одно банкротство?

4. На склад поступило 30 холодильников. Известно, что пять холодильников с дефектами, но неизвестно − какие. Найти вероятность того, что три взятых наугад холодильника будут с дефектами. Вероятность для компании, занимающейся строительством терминалов для аэропортов, получить контракт в стране *A* равна 0,4; получить в стране *B* равна 0,3. Чему равна вероятность того, что компания получить контракт хотя бы в одной стране?

Контрольная работа 1.

1. Три стрелка произвели залп по цели. Вероятность поражения цели для каждого стрелка равна соответственно 0,7; 0,8; 0,9. Найти вероятность того, что:

а) только один стрелок поразит цель;

б) только два стрелка поразят цель;

в) все три стрелка поразят цель;

г) хотя бы один стрелок поразит цель.

2. Стрельба производится по пяти мишеням типа А, трем – типа В, двум – типа С. Вероятность поражения мишени при одном выстреле типа А равна 0,4, типа В – 0,1, типа С – 0,15. Найти вероятность поражения мишени при одном выстреле, если известно, в мишень какого типа он сделан.

3. Вероятность выживания бактерий после радиоактивного облучения равна 0,004. Найти вероятность того, что после обучения 500 бактерий останется не менее 3 бактерий.

4. Вероятность наступления события в каждом из независимых испытаний равна 0,2. Сколько нужно произвести испытаний чтобы с вероятностью 0,9876 можно было ожидать отклонение относительной частоты появления события от его вероятности не более чем на 0,04?

5. Прибор состоит из пяти независимо работающих элементов. Вероятность отказа элемента, в момент включения прибора равна 0,2.
Найти: а) наивероятнейшее число отказавших элементов;

б) вероятность наивероятнейшего числа отказавших элементов;

в) вероятность отказа прибора, если для этого достаточно, чтобы отказали хотя бы 4 элемента.

6. По результатам проверок налоговыми инспекциями установлено, что в среднем каждое второе малое предприятие региона имеет нарушение финансовой дисциплины. Найти вероятность того, что из 1000 зарегистрированных в регионе малых предприятий имеют нарушения финансовой дисциплины: а) 480 предприятий; б) от 480 до 520. в) наивероятнейшее число предприятий число имеющих нарушение финансовой дисциплины.

7. Из статистики деятельности биржи известно, что в среднем цены на 75% котирующихся на бирже акций в течение месяца повысятся или останутся неизменными, а цены на 25% акций упадут. Некто имеет 6 акций, котирующихся на данной бирже. Найти вероятность того, что: а) более половины имеющихся у него акций не упадут в цене; б) по крайней мере, одна акция упадет в цене.

Контрольная работа 2.

1. Вероятность работы каждого из четырех комбайнов без поломок в течение определенного времени равна 0,9. Составить закон распределения случайной величины Х – числа комбайнов, работавших безотказно. Построить график распределения вероятностей. Найти математическое ожидание, дисперсию, среднее квадратическое отклонение случайной величины Х.

2. Вероятность того, что покупатель совершит покупку в магазине 0,4. Составить закон распределения случайной величины Х – числа покупателей, совершивших покупку, если магазин посетило 3 покупателя. Найти математическое ожидание, дисперсию, среднее квадратическое отклонение случайной величины Х. Построить график распределения вероятностей.

3. Непрерывная случайная величина Х задана интегральной функцией:

$$F(x) = \begin{cases} 0 & \text{ïðè} \quad x \le 0, \\ \dfrac{x^3}{125} & \text{ïðè} \quad 0 < x \le 5, \\ 1 & \text{ïðè} \quad x > 5. \end{cases}$$

Определить: а) вероятность попадания случайной величины в интервал (2; 3); б) математическое ожидание, дисперсию, среднее квадратическое отклонение случайной величины; в) функции распределения изобразить графически.

4. Урожайность озимой пшеницы по совокупности участков распределяется по нормальному закону с параметрами: а=50 ц/га, σ=10 ц/га. Определить: а) какой процент участков будет иметь урожайность свыше 40 ц/га; б) процент участков с урожайностью от 45 до 60 ц/га.

5. Испытываются два независимо работающих элемента. Длительность безотказной работы первого имеет показательное распределение $F_1(t) = 1 - e^{-0,1t}$, второго $F_2(t) = 1 - e^{-0,05t}$. Найти вероятность того, что за время длительностью 20 часов: а) оба элемента будут работать; б) откажет

только один элемент; в) откажет хотя бы один элемент; г) оба элемента откажут.

6. Количество электроэнергии, потребляемой поселком в течение суток, является случайной величиной, математическое ожидание которой равно 4 тыс. кВт. ч. Оценить вероятность того, что в ближайшие сутки потребление энергии: а) превысит 8 тыс. кВт. ч.; б) не превысит 6 тыс. кВт. ч.

Задания для индивидуальной работы 1

Вариант 0.

1. В команде 6 отличных стрелков и 4 хороших. Из команды наугад вызывают трех человек. Какова вероятность того, что все они – отличные стрелки?

2. Два велосипедиста имеют одинаковую вероятность приехать к указанному месту в любой момент в течение тридцати минут. Определить вероятность того, что время ожидания одним другого будет не более 5 мин?

3. Рабочий обслуживает одновременно 4 станка, причем, на первом вероятность нарушения нормальной работы в течение часа после проверки составляет 0,1, на втором – 0,15, на третьем – 0,2, на четвертом –0,25. Какова вероятность бесперебойной работы всех четырех станков на протяжении одного часа?

4. Два стрелка независимо друг от друга стреляют в цель. Вероятность попадания в цель первым стрелком – P_1=0,8, вторым – P_2 =0,7. Какова вероятность того, что: а) оба попадут в цель; б) хотя бы один попадет; в) попадет равно один?

5. Определить вероятность того, что партия из 100 изделий, среди которых пять бракованных, будет принята при испытании наудачу выбранной половины всей партии, если условиями приема допускается бракованных изделий не более одного из пятидесяти.

6. Имеются две урны. В первой урне 3 белых и 4 черных шара, во второй – 6 белых и 3 черных. Из первой урны во вторую перекладывают, не

глядя, пять шаров, после чего из второй урны вынимают шар. Какова вероятность того, что он окажется белым?

7. Оптовая база снабжает 10 магазинов, от каждого из которых может поступить заявка на очередной день с вероятностью 0,4. Найти наивероятнейшее число заявок в день и вероятность получения этого числа заявок.

8. Установлено, что в среднем 0,5% шариков, изготовленных для подшипников, оказываются бракованными. Определить вероятность того, что среди поступивших на контроль 10 000 шариков бракованными окажутся 40 штук.

9. Всхожесть хранящегося на складе зерна равна 80%. Отбираются первые попавшиеся 100 зерен. Требуется определить вероятность того, что среди них число всхожих от 68 до 90 штук.

10. При условиях предыдущей задачи найти вероятность того, что доля всхожих зерен будет отличаться от 0,8 по абсолютной величине не более, чем на 0,1.

11. В ящике 2 белых и 4 чёрных шара. Один за другим вынимаются все имеющиеся в нём шары. Найти вероятность того, что последний шар будет чёрным.

12. В партии товара, состоящей из 30 мужских пальто, находится 20 изделий местного производства. Товаровед наудачу выбирает 3 изделия. Какова вероятность того, что все 3 изделия окажутся: а) местного производства; б) не местного производства.

13. Статистика запросов кредитов в банке такова: 10% - государственные органы, 30% - другие банки, остальные - физические лица. Вероятности невозврата взятого кредита соответственно таковы: 0,01, 0,05 и 0,2. Найти вероятность невозврата очередного запроса на кредит. Начальнику кредитного отдела доложили, что получено сообщение о невозврате кредита, но в факсимильном сообщении имя клиента было неразборчиво. Какова вероятность, что данный кредит не возвращает какой-то банк?

14. Какова вероятность выпадения хотя бы двух шестёрок при трёх бросаниях игральной кости?

15. Сколько существует способов составления в случайном порядке списка из 7 кандидатов для выбора на руководящую должность? Какова вероятность того, что кандидаты будут расставлены в списке по возрасту (от меньшего к большему)?

16. Модельер, разрабатывающий новую коллекцию одежды к весеннему сезону, создает модели в зеленой, черной и красной цветовой гамме. Вероятность того, что зеленый цвет будет в моде весной, модельер оценивает в 0,3, что черный – в 0,2, а вероятность того, что будет моден красный цвет – в 0,15. Предполагая, что цвета выбираются независимо друг от друга, оцените вероятность того, что цветовое решение коллекции будет удачным хотя бы по одному из выбранных цветов.

17. Агент по недвижимости пытается продать участок земли под застройку. Он полагает, что участок будет продан в течение ближайших 6 месяцев с вероятностью 0,9, если экономическая ситуация в регионе не будет ухудшаться. Если же экономическая ситуация будет ухудшаться, то вероятность продать участок составит 0,5. Экономист, консультирующий агента, полагает, что с вероятностью, равной 0,7, экономическая ситуация в регионе в течение следующих 6 месяцев будет ухудшаться. Чему равна вероятность того, что участок будет продан в течение ближайших 6 месяцев.

Задания для индивидуальной работы 2

Вариант 0.

1. Задан ряд распределения случайной величины *X*.

Найти: а) M 2. Задана *F(x)* – функция распределения случайной величины *X*.

1.

x_i	23	25	28	29
p_i	0,3	0,2	0,4	0,1

2. Задана *F(x)* – функция распределения случайной величины *X*.

Найти: а) f(x); б) M(X), D(X), σ(x); в) построить графики *F(x)* и *f(x)*.

$$2.\quad F(x)=\begin{cases} 0 & при\ \ x\le 0,\\ \left(\dfrac{x}{2}\right)^2 & при\ 0<x\le 2,\\ 1 & при\ x>2\ . \end{cases}$$

3. Задана *f(x)* – плотность вероятности случайной величины *X*.
Найти: а) M(X), D(X), σ(x); б) F(x); в) построить графики *F(x)* и *f(x)*.

$$3.\quad f(x)=\begin{cases} 0 & при\ \ x\le 0,\\ \dfrac{1}{4} & при\ \ 0<x\le 4,\\ 0 & при\ \ x>4. \end{cases}$$

4. Устройство состоит из трех независимо работающих элементов. Вероятность отказа каждого элемента в одном опыте равна 0,1. Составить закон распределения числа отказавших элементов в одном опыте, найти числовые характеристики.

5. Случайная величина *X*, сосредоточенная на интервале [*1,3*], задана квадратичной функцией $F(x)=ax^2+bx+c$, имеющий максимум при 3. Найти параметры *a*, *b*, *c*, плотность вероятностей *f(x)* и вычислить вероятность попадания случайной величины *X* в интервал [*1*; *2*]. Построить графики функций *f(x)* и *F(x)*.

6. Математическое ожидание и дисперсия случайной величины *X* соответственно равны *M(X)=8; D(X)=2,4*. Найти математическое ожидание и дисперсию случайных величин: а) $3x+2$; б) *5x*; в) *2x-5*.

7. Детали, выпускаемые цехом, по размеру диаметра распределены по нормальному закону. Стандартная длина диаметра детали равна *20*, среднее квадратическое отклонение 3. Найти: а) вероятность, что диаметр наудачу взятой детали будет больше 17 и меньше 26; б) вероятность, что диаметр детали отклонится от стандартной длины не более, чем на 1,5.

8. Точка брошена наудачу внутрь круга радиуса R. Вероятность попадания точки в любую область, расположенную внутри круга, пропорциональна площади этой области. Найти функцию распределения расстояния точки до центра круга.

9. Найти вероятность того, что среди 200 изделий окажется более трех бракованных, если в среднем бракованные изделия составляют 2%.

10. Автобусы некоторого маршрута идут строго по расписанию. Интервал движения 7 мин. Найти вероятность того, что пассажир, подошедший к остановке, будет ожидать автобус менее 3 мин.

11. Дискретная случайная величина X задана законом распределения.

x_i	0,1	0,4	0,6
p_i	0,2	0,3	0,5

Пользуясь неравенством Чебышева, оценить вероятность того, что $|X - M(X)| < \sqrt{0,4}$.

12. На поле площадью в 1000 га берется на выборку по 1 м^2 с каждого га и подсчитывается урожайность. Оценить вероятность того, что средняя выборочная урожайность будет отличаться от средней урожайности по всей площади не более чем на 0,3 ц, если дисперсия на каждом га не превышает 4.

13. Вероятность наличия зазубрин на металлических брусках, заготовленных для обточки, равна 0,2. Оценить вероятность того, что в партии из 1000 брусков отклонение числа пригодных брусков от 800 не превышает 4 %.

14. Квантиль уровня 0,25 нормально распределённой случайной величины X равен 30, а квантиль уровня 0,5 равен 80. Найти математическое ожидание и среднее квадратическое отклонение случайной величины.

15. В страховом обществе на год застраховано 2000 автолюбителей. В случае аварии страховое общество выплачивает автолюбителю 500 рублей. Какую минимальную стоимость страхового взноса следует установить, чтобы вероятность того, что страховое общество к концу года окажется в убытке была не дольше 0,0014 если вероятность автолюбителю попасть в

аварию равна 0,005.

16. Пусть T (час) – время, необходимое для ремонта грузового автомобиля, удовлетворяет экспоненциальному распределению с параметром 0,1 . Какова вероятность того, что время ремонта одного автомобиля не превышает 6 часов, и сколько часов в среднем затрачивается на ремонт одного автомобиля.

17. Дневная добыча угля в некоторой шахте распределена по нормальному закону с математическим ожиданием 760 т и стандартным отклонением 62 т. Найдите вероятность того, что в определённый день будут добыты, по крайней мере 800 т угля. Определите долю рабочих дней, в которых будет добыто от 720 до 830 т угля. Найдите вероятность того, что в данный день добыча угля окажется ниже 750 т.

18. Компания А покупает у компании В детали к контрольным приборам. Каждая деталь имеет точно установленное значение размера. Деталь, размер которой отличается от установленного размера более чем на 0,25 мм считается дефектной. Компания А требует от компании В, чтобы доля брака не превышала 1% деталей. Если компания В выполняет требование компании А, то каким должно быть допустимое максимальное стандартное отклонение размеров деталей? Учесть, что размер деталей есть случайная величина, распределённая по нормальному закону.

19. Рассмотрим несколько операций (Q_1, Q_2, Q_3) со случайным доходом. Вычислите для всех операций ожидаемый доход $\overline{Q}$, СКО г. Нанесите эти характеристики на единый рисунок, получив графическое изображение операций. С помощью взвешивающей формулы E($\overline{Q}$, г) = 5Q – г найдите лучшую и худшую операции.

Q_1	10	5	12	20
	0,1	0,2	0,5	0,2
Q_2	8	5	10	20

	0,3	0,2	0,1	0,4
Q_3	6	8	10	20
	0,3	0,1	0,2	0,4

20. Задано распределение двумерной случайной величины. Найти корреляционный момент и коэффициент корреляции двух случайных величин X и Y.

y \ x	1	2	3
1	0,2	0,25	0,15
2	0,14	0,06	0.2

Тестовые задания для контроля знаний.

1. Вероятность случайного события изменяется в интервале:

   а) [0; 1] б) [-1; 0] в) [-1; +1].

2. Классическое определение вероятности применимо, когда исходы испытания:

   а) равновозможны;

   б) не совместны;

   в) единственно возможны;

   г) практически невозможным

3. Если p(A)≈1, то события A называется:

   а) достоверным;

   б) невозможным;

   в) практически достоверным;

   г) практически невозможным;

8 .Если p(A) ≈ 0, то событие A называется:

а) достоверным;

б) невозможным;

в) практически невозможным;

г) практически достоверным;

9. Если события А и В равны, А={3; 5; 8}, В={4; 5; 7}, то событие А + В равно:

   а) {3; 4; 5; 7; 8}, б) {3; 8}, в) {5}.

10. Если события А и В равны, А={3; 5; 8}, В={4; 5; 7}, то событие А-В равно:

   а) {3; 4; 5; 7; 8}, б) {3; 8}, в) {5}.

11. Если события А и В равны, А={3; 5; 8}, В={4; 5; 7}, то событие А* В равно:

   а) {3; 4; 5; 7; 8}, б) {3; 8}, в) {5}.

12. Если р(А) = 07, то р($\overline{A}$) равно:

   а) 0; б)1; в) –0,7; г) 0,3.

13. Вероятность события, противоположного событию А, равна:

   а) 0; б)-р(А); в) 1-р(А); г) 1.

14. Согласно принципа практической уверенности практически достоверное событие:

   а) обязательно произойдет;

   б) обязательно не произойдет;

   в) может произойти или не произойдет.

15. Согласно принципа практической уверенности практически невозможное событие:

   а) обязательно произойдет;

   б) обязательно не произойдет;

   в) может произойти или не произойдет.

16. Для совместных событий А и В имеет место утверждение:

   а) р(А·В) = р(А)·р(В/А);

   б) р(А·В) = р(А)·р(В);

   в) р(А+В) = р(А)+р(В);

г) p(A+B) = p(A)+p(B)-p(A B).

17.Для несовместных событий A и B имеет место утверждение:

а) p(A·B) = p(A)·p(B/A);

б) p(A·B) = p(A)·p(B);

в) p(A+B) = p(A)+p(B);

г) p(A+B) = p(A)+p(B)-p(A B).

18.Для зависимых событий A и B имеет место утверждение:

а) p(A·B) = p(A)·p(B/A);

б) p(A·B) = p(A)·p(B);

в) p(A+B) = p(A)+p(B);

г) p(A+B) = p(A)+p(B)-p(A B).

19.Для независимых событий A и B имеет место утверждение:

а) p(A·B) = p(A)·p(B/A); б) p(A·B) = p(A)·p(B);

в) p(A+B) = p(A)+p(B); г) p(A+B) = p(A)+p(B)-p(A B).

20.Если событие A,B,C образуют полную группу несовместных событий и p(A)=0,3, p(B)=0,6, то вероятность события C равна:

а) 0,9; б)0,18; в) 0,1; г) 0,2.

21.Если p(A) = 0,24, p(A·B)=0,06, то p(B/A) равна:

а) 0,26; б)0,14; в) 0,3; г) –0,14.

22.Если событие A не зависит от B, то :

а) p(A)=p(A/B); б)p(B)=P(A/B); в) p(A/B)=P(B/A).

23.Событие, состоящие из всех элементарных событий, принадлежащих по крайней мере одному из двух событий есть:

а) сумма этих событий;

б) произведение этих событий;

в) разность этих событий.

24. Событие, состоящие из всех элементарных событий, принадлежащих каждому из двух событий есть:

а) сумма этих событий;

б) произведение этих событий;

в) разность этих событий.

25. Частный случай схемы независимых испытаний, в котором каждое испытание может закончиться только одним из двух исходов, назвают:

а) полиномиальной схемой;

б) конечной вероятностной схемой;

в) схемой Бернулли.

26. Формула интерпретируемая как формула, позволяющая по априорным вероятностям и по результатом опыта найти апостериорные вероятности есть:

а) формула Пуассона;

б) формула Байеса;

в) формула Лапласа.

27. События $A_1, A_2, ... A_n$ образуют полную группу событий, если в результате испытания:

а) произойдут хотя бы два из них;

б) одно из них обязательно произойдет;

в) все события $A_1, A_2, ... A_n$ обязательно произойдут.

28. Число, характеризующие возможность появления события, называется:

а) алгеброй событий;

б) элементарным исходом;

в) вероятностью события.

29. Символом C_n^m обозначается:

а) число перестановок из n элементов по n;

б) число сочетаний из n элементов по m;

г) число размещений из n элементов по m.

30. Символом A_n^m обозначается:

а) число перестановок из n элементов по n;

б) число сочетаний из n элементов по m;

г) число размещений из n элементов по m.

31. Символом $P_n(n = m)$ обозначается:

а) число перестановок из n элементов по n;

б) число сочетаний из n элементов по m;

г) число размещений из n элементов по m.

32. Размещения из n элементов по m отличаются друг от друга:

а) порядком элементов;

б) хотя бы одним элементом;

в) либо порядком, либо самими элементами.

33. Сочетания из n элементов по m отличается друг от друга:

а) порядком элементов;

б) хотя бы одним элементом;

в) либо порядком, либо самими элементами.

34. Теорема сложения несовместных (совместных) событий утверждает, что:

а) p(A+B) = p(A) + p(B);

б) p(A+B) = p(A) + p(B) - p(A·B);

в) p(A+B) = p(A) + p(B) + p(A·B)

35. Формула Пуассона $P_{m,n} \approx \frac{\lambda}{m!} e^{-\lambda}$, где $\lambda = np$, применяется, если:

а) n-достаточно велико, а p- достаточно мало;

б) n- достаточно мало а p- достаточно велико;

в) n и p- достаточно велики;

г) n и p- достаточно малы.

36. Точное значение вероятности $P_{m,n}$ находится:

а) по формуле Бернулли;

б) по формуле Пуассона;

в) по формуле Лапласа;

г) по формуле Бейеса;

д) по формуле полной вероятности.

37. Если A $\subset$ B, то:

а) p(A) = p(B);

б) $p(A) \geq p(B)$;

в) $p(A) \leq p(B)$.

38. Формула $P_{m,n} = C_n^m \ p^m \ q^{n-m}$, где q = 1-p, есть формула:

а) Бернулли; б) Пуассона; в) Бейеса.

39. Вероятность $P_{m,n}$ при больших значениях n и p находятся по:

а) формуле Бернулли;

б) формуле Пуассона;

в) локальной теореме Лапласа.

40. Число, около которого группируются частости события по мере увеличения числа испытаний, есть:

а) вероятность события;

б) наивероятнейшее число;

в) принцип практической уверенности.

41. Наивероятнейшее число – это:

а) число, около которого группируются частности данного события по мере увеличения числа испытаний;

б) число, характеризующее возможность появления события;

в) число, для которого вероятность появления события в испытаниях, наибольшая;

г) все ответы неправильны.

42. Число, характеризующее возможность появления события, – это:

а) наивероятнейшее число;

б) противоположное событие;

в) вероятность события;

г) теория вероятности;

д) принцип практической уверенности.

43. События независимые, если:

а) безусловная вероятность равна условной вероятности;

б) безусловная вероятность не равна условной вероятности;

в) безусловная вероятность больше условной вероятности;

г) безусловная вероятность меньше условной вероятности;

44. Произведение двух событий, если они зависимые:

а) p(A·B)=p(A)·p(B);

b) б) p(A·B)=p(A)·p(B/A);

в) p(A·B)=p(B)·p(A/B);

г) правильные ответы а и б;

д) правильные ответы б и в;

е) правильные ответы а, б, в.

45. Множество взаимосвязанных исходов, таких, что в результате эксперимента появляется один и только один из эти исходов, называется:

а) пространством элементарных исходов;

б) алгеброй событий;

в) вероятностью появления события.

46. Формула вида $P(A)=\sum_{i=1}^{n}P(H_i)P(A/H_i)$ называется:

а) формулой Лапласа;

б) формулой Байеса;

в) полной вероятностью события А.

47. Принцип, состоящий в том, что если событие практически достоверное, то считаем, что оно обязательно произойдет, называется принципом:

а) практической возможности;

б) практической уверенности;

в) практической достоверности.

48. Если вероятность появления одного из событий не зависит от того произойдет или не произойдет другое событие, то эти два события называются:

а) зависимыми;

б) совместными;

в) несовместными.

49. Если вероятность появления события не равна единицы, то близка к ней, то событие называется:

а) практически достоверными;

б) практически невозможными;

в) достоверными.

50. Формула поной вероятности имеет вид:

а) $P(H_i / A) = \dfrac{P(H_i) \times P(A/H_i)}{\sum P(H_i) \cdot P(A/H_i)}$;

б) $P(A) = \sum P(H_i) \cdot P(A/H_i)$;

в) $P(A) = m/n$.

51. Формула Бернулли имеет вид:

а) $P_{m,n} = C_n^m \; p^m q^{n-m}$;

б) $P_{m,n} = \dfrac{\lambda^m}{m!} e^{-\lambda}$; где λ = пр;

в) $P_{m,n} = \dfrac{\varphi(x)}{\sqrt{npq}}$; где $\varphi = \dfrac{1}{\sqrt{2\overline{n}}} \cdot e^{\frac{x^2}{2}}$, $x = \dfrac{m - np}{\sqrt{npq}}$.

52. Закон редких явлений – это:

а) Закон Пуассона;

б) локальная теорема Лапласа;

в) интегральная теорема Лапласа.

53. Если вероятность наступления события А постоянна, а n- независимых испытаний достаточно велико, то $P_{m,n}$ находятся с помощью:

а) локальной теорем Лапласа;

б) теорем Пуассона;

в) интегральной теорем Лапласа.

54. Вероятность отклонения относительной частоты события от вероятности в независимых испытании находится из соотношения:

а) пр-q $\le m_0 \le np + p$;

б) пр-n$\varepsilon < m < np + n\varepsilon$;

в) $P(\left|\frac{m}{n} - p\right| \le \text{E}) = 2\phi(\varepsilon\sqrt{\frac{n}{pq}})$.

55. По формуле $C_n^m = \frac{n!}{m!(n-m)!}$, находится:

а) число размещений из n элементов по m;

б) число сочетаний из n элементов по m;

в) число перестановок из n элементов.

56. По формуле $A_n^m = \frac{n!}{(n-m)!}$, находится:

а) число размещений из n элементов по m;

б) число сочетаний из n элементов по m;

в) число перестановок из n элементов.

57. Если U- пространство элементарных исходов, то имеет равенство:

а) $\overline{A} = U - A$; б) $A = U - \overline{A}$; в) $U = A + \overline{A}$; г) все равенства верны.

58. Игральная кость бросается один раз. Тогда вероятность того, что на верхней грани выпадет ***не более пяти очков***, равна…

1) $\frac{1}{6}$	2) $\frac{2}{3}$	3) $\frac{5}{6}$	4) 1

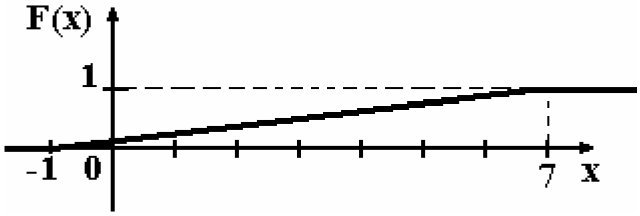
59. Дискретная случайная величина ***X*** задана законом распределения вероятностей:

X	-1	0	3
p	0,1	0,3	0,6

Тогда математическое ожидание случайной величины ***Y=3X*** равно…

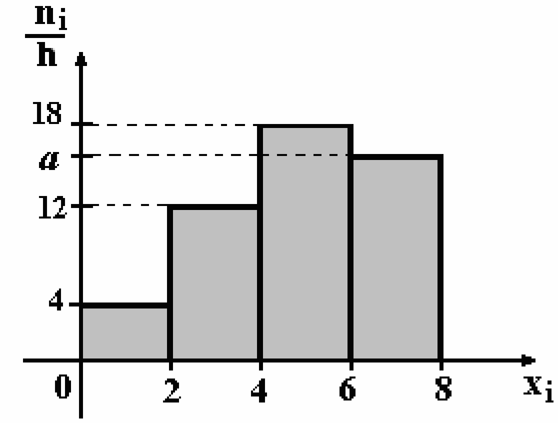
1) 5,7	2) 6	3) 5,1	4) 4,7

60. График функции распределения вероятностей непрерывной случайной величины ***X*** имеет вид:

 	Тогда математическое ожидание ***X*** равно…

1) 7	2) 8	3) 4	4) 3

61. По выборке объема n=100 построена гистограмма частот:

 	Тогда значение ***a*** равно…

1) 16	2) 66	3) 15	4) 17

62. В результате измерений некоторой физической величины одним прибором (без систематических ошибок) получены следующие результаты (в

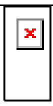
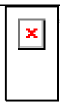
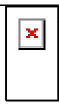
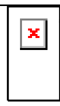
1) 13	2) 2	3) 6	4) 3

мм): 11, 14, 14. Тогда несмещенная оценка дисперсии измерений равна…

63. По оценкам экспертов вероятности банкротства для двух предприятий, производящих однотипную продукцию, равны 0,1 и 0,15. Тогда вероятность банкротства обоих предприятий равна…

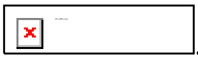
1) 0,765 2) 0,15 3) 0,25 4)0,015

64. В первой урне 3 белых и 7 черных шаров. Во второй урне 5 белых и 15 черных шаров. Из наудачу взятой урны вынули один шар. Тогда вероятность того, что этот шар окажется белым, равна…

1) 	2) 	3) 	4) 

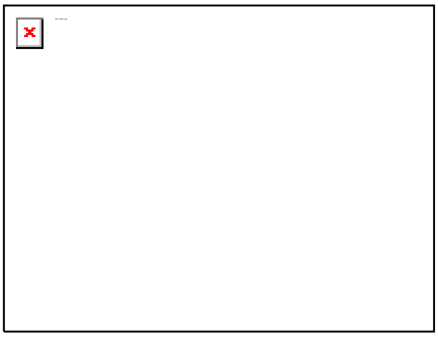
65. Из генеральной совокупности извлечена выборка объема n=50:

Тогда n_4 равен…

66. Если основная гипотеза имеет вид , то конкурирующей может быть гипотеза…

1) 	2) 	3) 	4) 

67. Из генеральной совокупности извлечена выборка объема n=50, полигон частот которой имеет вид



Тогда число вариант x_i=4 в выборке равно…

1) 16	2) 14	3) 50	4) 15

68. Мода вариационного ряда ***1, 2, 2, 3, 4, 5*** равна…

1) 3	2) 5	3) 2	4) 17

69. По оценкам экспертов вероятности банкротства для двух предприятий, производящих однотипную продукцию, равны 0,1 и 0,15. Тогда вероятность банкротства обоих предприятий равна…

1) 0,25	2) 0,765	3) 0,15	4) 0,015

70. Вероятность появления события А в 10 независимых испытаниях, проводимых по схеме Бернулли, равна 0,8. Тогда дисперсия числа появлений этого события равна…

1) 1,6	2) 0,08	3) 8	4) 0,16

Экзаменационные вопросы

1. Случайные события и их классификация.
2. Элементы комбинаторики.
3. Различные подходы к введению вероятности. Практически невозможные и практически достоверные события. Принцип практической уверенности.
4. Теорема сложения вероятностей несовместимых событий.
5. Теорема умножения вероятностей.
6. Теорема сложения вероятностей, совместимых событий.
7. Формула полной вероятности.
8. Теорема Байеса.
9. Формула Бернулли. Наивероятнейшее число наступления события.
10. Теоремы Лапласа.
11. Вероятность отклонения частоты от постоянной вероятности в независимых испытаниях.
12. Случайные величины. Закон распределения дискретных случайных величин.
13. Функция распределения и ее свойства.
14. Плотность распределения и ее свойства.
15. Функция случайной величины и ее распределение.
16. Математическое ожидание случайной величины, свойство математического ожидания.
17. Дисперсия случайной величины, свойства дисперсии.

18. Биноминальное распределение.
19. Распределение Пуассона.
20. Равномерное распределение.
21. Нормальное распределение.
22. Вероятность попадания нормального распределенной величины на заданный участок. Правило трех сигм.
23. Понятие о теореме Ляпунова.
24. Оценка отклонения распределения случайной величины от нормального.
25. Эксцесс и асимметрия.
26. Показательное распределение.
27. Неравенство Чебышева.
28. Теорема Чебышева.
29. Теорема Бернулли.
30. распределения системы случайных величин, таблица распределения.
31. Функция распределения системы случайных величин, ее свойства.
32. Плотность распределения системы случайных величин, ее свойства.
33. Плотность распределения отдельных величин, входящих в систему. Условные законы распределения.
34. Зависимые и независимые случайные величины.
35. Числовые характеристики системы случайных величин.
36. Понятие о случайном процессе и его реализации.
37. Марковский случайный процесс.
38. Понятие выборки случайных величин.
39. Понятие о выборочном методе.
40. Понятие генеральной совокупности (генеральной выборки).
41. Понятие регрессии, регрессионные зависимости.
42. Регрессионная зависимость как “ослабленная” функциональная зависимость.
43. Виды регрессионной зависимости.

44. Метод наименьших квадратов как метод аналитического сглаживания и определения параметров регрессионной зависимости.
45. Множественная регрессия.
46. Корреляционная зависимость между случайными величинами.
47. Ковариация. Коэффициент корреляции.
48. Различия между регрессионной и корреляционной зависимостями.
49. Основные задачи математической статистики.
50. Статистическая функция распределения.
51. Статистический ряд. Гистограмма.
52. Числовые характеристики статистического распределения.
53. Выравнивание статистических рядов, метод наибольшего правдоподобия.
54. Свойства точечных оценок.
55. Доверительный интервал. Доверительная вероятность.
56. Доверительный интервал для оценки математического ожидания нормального распределения.
57. Функциональная, статистическая и корреляционная зависимости.
58. Статистические гипотезы.
59. Статистический критерий проверки нулевой гипотезы. Наблюдаемое значение критерия.
60. Критическая область. Критические точки и их нахождение.
61. Сравнение двух дисперсий нормальных генеральных совокупностей.
62. Сравнение двух средних нормальных генеральных совокупностей, дисперсии которых известны. Понятие о критериях согласия.

Образцы экзаменационных билетов

ЭКЗАМЕНАЦИОННЫЙ БИЛЕТ № 1

1. Докажите теорему Чебышева.
2. Выборочная дисперсия. Оценка генеральной дисперсии по исправленной выборочной дисперсии.
3. Аксиомы теории вероятностей.
4. Решите задачи:

1) В урне 7 белых и 6 черных шаров. Случайным образом вынимают 5 шаров. Найти вероятность того, что среди них имеются 3 белых шара.
2) Вероятность выживания бактерий после радиоактивного облучения 0,002. Найти вероятность того, что после облучения 1000 бактерий выживут не менее 3-х бактерий.
3) $f(x)=\begin{cases} 0 & \text{ïðè} \quad x \le 0, \\ 3x^2 & \text{ïðè} \quad 0 < x \le 1, \\ 0, & \text{ïðè} \quad x > 1. \end{cases}$

Найти M(x), D(x), F(x), графики обоих функций.

4) При установившемся технологическом процессе 2/5 всей продукции автомат выпускает 1-м сортом и 3/5 – 2-м сортом. Составить закон распределения числа изделий первого сорта среди 4-х случайно отобранных изделий.
5) Начиная с какого числа n независимых испытаний имеет место неравенство $P\left\{\left|\frac{m}{n}-0{,}8\right|<0{,}04\right\}>0{,}92$.

ЭКЗАМЕНАЦИОННЫЙ БИЛЕТ № 2

1. Отыскание выборочного уравнения прямой линии регрессии по несгруппированным данным.
2. Алгоритм решения задач с примечанием теорем сложения и умножения вероятностей.
3. Значение теоремы Бернулли.
4. Решите задачи:
    1) Студент знает 20 из 30 вопросов программы. Найти вероятность того, что студент знает 2 вопроса экзаменационного билета, содержащего 3 вопроса.
    2) Вероятность появления события в каждом испытании 0,25. Найти вероятность того, что событие наступит 50 раз в 243 испытаниях.
    3) Составить ряд распределения х-числа появлений «герба» в 3-х независимых испытаниях. Найти M(x), D(x), F(x).
    4) Книга в 500 страниц содержит 50 опечаток. Оценить вероятность того, что на случайно выбранной странице не мене 3-х опечаток.
    5) Оценить вероятность неравенства: $\left|\frac{m}{n}-0{,}75\right|<0{,}05$.

ЭКЗАМЕНАЦИОННЫЙ БИЛЕТ № 3

1. Дайте понятие и сформулируйте и докажите свойства функции распределения случайных величин.
2. Установите сходства и различия в применении формул Бернулли, Пуассона и локальной теоремы Лапласа.

3. Дайте понятия и приведите примеры простой и сложной гипотез.
4. Решите задачи:
    1) Из трех станков, обслуживаемых рабочим, вероятность остановки на протяжении часа составляют для 1-0,3; 2-0,25; 3-0,4. Найти вероятность бесперебойной работы двух станков на протяжении часа.
    2) Из 20 стрелков 5 попадают в мишень с вероятностью 0,8; 6 – с вероятностью 0,7; 5 – с вероятностью 0,4; 4 – с вероятностью 0,5. Наудачу выбранный стрелок произвел выстрел, но в мишень не попал. К какой из групп вероятнее всего принадлежал этот стрелок?
    3) $f(x) = \begin{cases} 0 & \text{ïðè} \quad x \le 0, \\ x/2 & \text{ïðè} \quad 0 < x \le 2, \\ 0, & \text{ïðè} \quad x > 2. \end{cases}$

    Найти M(x), D(x), F(x).
    4) Автомат изготавливает шарики. Шарик считается годным, если отклонение диаметра шарика от проектного размера по абсолютной величине меньше 0,8 мм. Считая, что распределена нормально с параметрами $\sigma = 0{,}6$ мм, $a = 0$. Найти сколько будет годных шариков среди 200 изготовленных.
    5) Найти ε, если: $P\{|x - M(x)| < \varepsilon\} \ge 0{,}96$ и $D(x) = 0{,}03$.

ЭКЗАМЕНАЦИОННЫЙ БИЛЕТ № 4

1. Докажите теорему сложения вероятностей совместных событий. Приведите пример ее применения.
2. Сформулируйте алгоритм проверки статистических гипотез.
3. Перечислите законы распределения дискретных случайных величин.
4. Решите задачи:
    1) Бросают две монеты. Определить вероятность того, что хотя бы на одной монете появится герб.
    2) Два автомата производят детали, которые поступают на общий конвейер. Вероятность брака на первом автомате 0,06, на втором – 0,09. Производительность второго автомата вдвое больше, чем первого. Найти вероятность бракованной детали с конвейера.
    3) Дан ряд распределений:

x_i	5	8	11
p_i	0,2	0,4	0,4

    Найти M(x), D(x), F(x) и ее график.
    4) Случайные ошибки измерения подчинены нормальному закону с параметрами a=0, $\sigma = 20$ мм. Найти вероятность того, что из трех независимых измерений ошибка хотя бы одного не превзойдет 4 мм по абсолютной величине.

5) Вероятность наступления события А в каждом испытании 0,2. Оценить вероятность того, что в 10000 испытаниях отклонение частости события А от его вероятности не превзойдет по абсолютной величине 0,01.

Литература

Основная

1. Красс М.С., Чупрынов Б. П. "Математика для экономических специальностей" - Москва : ИнФРА-М,2004.

2. Кремер И.Ш., Теория вероятностей и математическая статистика. – М.: Юнити, 2006. – 573 с.

3. Гмурман В.Е. Руководство к решению задач по теории вероятностей и математической статистике. - М.: Высшая школа, 2005. – 404 с.

4. Гмурман В.Е. . Теория вероятностей и математическая статистика. – М.: Высшая школа, 2004. – 479 с.

Дополнительная

5. Колемаев В.А. , Калинина В.Н. Теория вероятности и математическая статистика.- Москва.: ИнФРА-М,1997.
6. Горелова Г. В., Кацко И. А. Теория вероятности и математическая статистика.- Ростов-на- Дону.: Феникс, 2002.

Методические разработки кафедры.

7. Торопчина Г. Н., Вохминцева Г. П., Шевченко И. Н. Введение в теорию вероятностей. - Благовещенск.: АмГУ, 2006.
8. Торопчина Г. Н., Вохминцева Г. П., Шевченко И. Н. Лабораторные работы по математической статистике. - Благовещенск.: АмГУ, 2006.

Содержание

1 Пояснительная записка 3

2 Государственный стандарт 4

3 Содержание курса 5

4 Темы лекций 8

5 Календарный план занятий 10

6 Конспект лекций по теории вероятностей 12

7 Содержание практических занятий 102

8 Методические указание к практическим занятиям 102

9 Методические указания для выполнения домашних и контрольных работ 126

10 Примерные задания для контрольных и расчетно-графических работ 130

11 Тестовые задания для контроля знаний 139

12 Экзаменационные вопросы 150

13 Образцы экзаменационных билетов 152

Литература 155