ТЕОРИЯ ВЕРОЯТНОСТИ И МАТЕМАТИЧЕСКАЯ СТАТИСТИКА

Лекция 6.

Основные изучаемые вопросы:

- Закон больших чисел.
- Лемма Маркова.
- Неравенство Чебышева.
- Теорема Чебышева.
- Теорема Бернулли.
- Теорема Пуассона.
- Центральная предельная теорема.

ПРЕДЕЛЬНЫЕ ТЕОРЕМЫ ТЕОРИИ ВЕРОЯТНОСТЕЙ И ЗАКОН БОЛЬШИХ ЧИСЕЛ

- ***В широком смысле*** под законом больших чисел понимается ***свойство устойчивости массовых явлений***, состоящее в том, что ***средний результат действия большого числа случайных явлений практически перестает быть случайным и может быть предсказан с достаточной определенностью***.
- ***В узком смысле*** под законом больших чисел понимают ***совокупность теорем, устанавливающих факт приближения средних характеристик, полученных по результатам большого числа наблюдений, к некоторым постоянным величинам.***

Лемма Маркова

- Если случайная величина X не принимает отрицательных значений, то для любого положительного числа t выполняется соотношение

$$P(X \geq t) < \frac{M[X]}{t}.$$

- Пусть X – непрерывная случайная величина с плотностью $f(x)$.
- Математическое ожидание случайной величины X равно

$$M[X] = \int_0^\infty x \cdot f(x) \cdot dx > \int_t^\infty x \cdot f(x) \cdot dx.$$

- Последнее неравенство записано исходя из того, что $x > 0$ и $f(x) > 0$.
- Так как $x > t$, то

$$M[X] > \int_t^\infty t \cdot f(x) \cdot dx = t \int_t^\infty f(x) \cdot dx = t \cdot P(X \geq t).$$

- Отсюда следует соотношение приведенное в верхней части слайда.

- **Пример.** Среднее число вызовов наладчика станков за смену равно 21. Оценить вероятность того, что за смену поступит вызовов:

  - не менее 60;

  - менее 35.

  Число вызовов – положительная случайная величина X с $M[X] = 21$.

  В соответствии с леммой Маркова для первого случая $t = 60$:

$$P(X \geq 60) < \frac{21}{60} = 0{,}35.$$

  Для второго случая ($t = 35$) вероятность ищем через обратное событие:

$$P(X < 35) = 1 - P(X \geq 35) > 1 - \frac{21}{35} = 1 - 0{,}6 = 0{,}4.$$

- **Пример.** Пусть X – число очков, выпавших на игральной кости. Оценить вероятность того, что X примет значение, меньшее 5, пользуясь леммой Маркова.
- Математическое ожидание числа выпавших очков есть

$$M[X]=1\cdot\frac{1}{6}+2\cdot\frac{1}{6}+3\cdot\frac{1}{6}+4\cdot\frac{1}{6}+5\cdot\frac{1}{6}+6\cdot\frac{1}{6}=3{,}5.$$

- Тогда в соответствии с леммой Маркова при t = 5 имеем:

$$P(X<5)=1-P(X\geq 5)>1-\frac{3{,}5}{5}=0{,}3.$$

- Полученный результат не противоречит вычисленной нами ранее вероятности с помощью классического определения

$$P(X<5)=\frac{n}{m}=\frac{4}{6}=0{,}667.$$

- Действительно: 0,667 > 0,3.

Неравенство Чебышева

- Определение. ***Для любой случайной величины X, имеющей конечную дисперсию, при каждом ε > 0 имеет место неравенство:***

$$P(|X - M(X)| \le \varepsilon) \ \ge \ 1 - \frac{D(X)}{\varepsilon^2}.$$

- Или альтернативная форма (через обратное событие):

$$P(|X - M(X)| \ge \varepsilon) \ \le \ \frac{D(X)}{\varepsilon^2}.$$

- ***Вероятность того, что реализация величины X отклонится от своего математического ожидания не менее чем на ε, ограничена сверху величиной D/ε^2.***

- Пусть $\varepsilon = t \cdot \sigma$, а σ – среднеквадратическое отклонение величины X. Тогда можно записать:

$$P(|X - M(X)| \ge t \cdot \sigma) \ \le \ \frac{D(X)}{t^2 \cdot \sigma^2} = \frac{1}{t^2}.$$

- Оценим вероятности того, что случайная величина отклонится от математического ожидания на величину σ, 2σ и 3σ.

$$при\ t=1 \qquad P(|X-M(X)|\geq\sigma)\ \leq\ 1;$$

$$при\ t=2 \qquad P(|X-M(X)|\geq 2\sigma)\ \leq\ \frac{1}{4};$$

$$при\ t=3 \qquad P(|X-M(X)|\geq 3\sigma)\ \leq\ \frac{1}{9}.$$

- Из последнего неравенства следует, что вероятность отклонения значения случайной величины *X* от среднего значения не менее (1 – 1/9) = 8/9 ***при любом законе распределения случайной величины***.
- Для большинства встречающихся на практике законов распределения эта величина намного больше (например, для нормального закона она равна 0,997…). ***Однако этот факт не противоречит неравенству Чебышева. Оно дает нижнюю границу оценки вероятности.***

Пример. Электрическая подстанция обслуживает сеть с 10 000 ламп, вероятности включения каждой из которых вечером равна 0,6. Оценить вероятность того, что число одновременно включенных ламп будет находиться в пределах от 5900 до 6100.

- ***Решение.***
- Случайная величина X - число одновременно включенных ламп. Так как вероятность включения каждой из лампочек есть величина постоянная, и при этом включение каждой лампочки является независимым событием, то X распределена по биномиальному закону. Поэтому ее числовые характеристики:

$$M(X) = n \cdot p = 10\,000 \cdot 0{,}6 = 6000,$$

$$D(X) = n \cdot p \cdot q = 10\,000 \cdot 0{,}6 \cdot 0{,}4 = 2400.$$

- Событие, состоящее в том, что случайная величина X будет находиться в пределах от 5900 до 6100 включительно, означает:

$$5900 < X < 6100;$$

$$5900 - 6000 < X - 6000 < 6100 - 6000;$$

$$-100 < X - 6000 < 100,$$

т. е. $|X - 6000| < 100$.

- Так как для случайной величины X существует математическое ожидание и ограниченная дисперсия, то в соответствии с неравенством Чебышева:

$$P(|X - M(X)| \le \varepsilon) \ge 1 - \frac{D(X)}{\varepsilon^2}.$$

- По данным задачи получаем:

$$P(|X - 6000| < 100) \ge 1 - 2400/100^2 = 0,76.$$

- Значит, вероятность того, что число одновременно включенных ламп будет находиться в пределах от 5900 до 6100 включительно, - не менее 0,76.

Теорема Чебышева

- Если X_1, X_2,..., X_n, ... - последовательность независимых случайных величин, имеющих конечные дисперсии, ограниченные одной и той же постоянной C, то каково бы ни было $\varepsilon > 0$:

$$\lim_{n\to\infty} P\left\{\left|\frac{1}{n}\sum_{i=1}^{n} X_i - \frac{1}{n}\sum_{i=1}^{n} M(X_i)\right| < \varepsilon\right\} = 1.$$

- ***При достаточно большом числе независимых опытов среднее арифметическое значений случайной величины X сходится по вероятности к ее математическому ожиданию.***
- Иными словами, при большом числе опытов среднее значение случайной величины и ее математическое ожидание будут сколь угодно близки.
- Тем самым оправдывается рекомендуемый в практической деятельности способ получения более точных результатов измерений: одна и та же величина измеряется многократно, и в качестве ее значения ***берется среднее арифметическое полученных результатов измерений***.

- ***Теорема Чебышева имеет громадное практическое значение.*** Так, если производится измерение какой-либо величины и измерительный прибор не дает систематической ошибки, то результаты X_1 X_2, ..., X_n отдельных измерений являются случайными величинами, математические ожидания которых равны истинному значению $M[X]$ измеряемой величины. Дисперсии $D[X_1] < C$, $D[X_2]$ ... $< C$ ограничены, так как точность прибора, как правило, известна.
- Из теоремы Чебышева следует, что ***можно достоверно получить меньшую ошибку, если произвести не одно, а несколько измерений и за истинное значение измеряемой величины принять среднее арифметическое результатов отдельных измерений.***
- Именно так и поступают на практике, когда требуется достичь большей точности для значения измеряемой величины.

- ***Закон больших чисел действует во многих физических явлениях.***
- **Пример**. Газ состоит из множества беспорядочно движущихся молекул. Предсказать поведение одной молекулы невозможно, но совокупное их действие подчиняется закону больших чисел. Например, ***давление газа, которое определяется суммарным воздействием молекул на единицу площади, должно быть практически постоянным, что и наблюдается в действительности.***
- ***Теорема Чебышева служит обоснованием выборочного метода***, широко применяемого в статистике, когда по случайной выборке судят обо всей совокупности исследуемых объектов. Например, оценивают качество партии конфет по сравнительно небольшой пробе, считая, что проба содержит все же достаточно много конфет для того, чтобы проявлялось действие закона больших чисел.

Теорема Бернулли

- Пусть m - число наступления события A в серии n независимых испытаний, а p - вероятность наступления события в каждом из испытаний. Тогда каково бы ни было $\varepsilon > 0$:

$$\lim_{n\to\infty} P\left\{\left|\frac{m}{n} - p\right| < \varepsilon\right\} = 1.$$

- С увеличением числа опытов частота события A ***стремится к вероятности этого события.***
- Применим неравенство Чебышева к случайной величине m/n. Матожидание и дисперсию определим так:

$$M\left[\frac{m}{n}\right] = \frac{1}{n} M[m] = \frac{1}{n} np = p,$$

$$D\left[\frac{m}{n}\right] = \frac{1}{n^2} D[m] = \frac{1}{n^2} npq = \frac{pq}{n}.$$

- Тогда в соответствии с неравенством Чебышева

$$P\left\{\left|\frac{m}{n} - p\right| < \varepsilon\right\} \geq 1 - \frac{pq}{n\varepsilon^2}.$$

- В каждой области человеческой деятельности, как правило, существует определенный «уровень значимости» вероятностей, т.е. такое граничное число *p*, что события, имеющие вероятность, меньшую *p*, считаются практически невозможными.
- Если в качестве примера взять партию пылесосов, то если уровень значимости (вероятность бракованного изделия) будет находиться в пределах 0,01-0,001, этот уровень будет приемлем для покупателя, так как в данном случае вероятность порядка 0,001 достаточно мала, чтобы рассматривать бракованный пылесос как практически невозможное явление. Однако та же самая вероятность 0,001 велика, если речь идет об изготовлении, например, продуктов питания.

Теорема Пуассона

- Пусть m - число наступления события A в серии n независимых испытаний, а p_i - вероятность наступления события в i-м испытании. Тогда каково бы ни было $\varepsilon > 0$:

$$\lim_{n\to\infty} P\left\{\left|\frac{m}{n} - \frac{1}{n}\sum_{i=1}^{n} p_i\right| < \varepsilon\right\} = 1.$$

- ***При неограниченном возрастании числа независимых опытов, <u>производимых в изменяющихся условиях</u>, в каждом из которых событие А может появиться с вероятностью p_i, частота события А (m/n) сходится по вероятности к среднему арифметическому вероятностей p_i.***

Центральная предельная теорема

- Выше были рассмотрены различные формы ***закона больших чисел***, которые, при всем своем разнообразии, утверждают одно: ***сходимость по вероятности тех или иных случайных величин к определенным постоянным.***
- Все формы центральной предельной теоремы посвящены другому аспекту - ***установлению условий, при которых возникает самый распространенный в случайных явлениях нормальный закон распределения.*** Важнейшее место занимает теорема Ляпунова.

Теорема Ляпунова

- Рассмотрим n независимых случайных величин $X_1, X_2, \ldots, X_n$, удовлетворяющих условиям:

  1) все величины имеют определенные математические ожидания $M[X_i]$ и конечные дисперсии $D[X_i]$;

  2) ни одна из величин не выделяется резко среди остальных по своим значениям.

- ***Тогда при неограниченном возрастании n распределение случайной величины $\frac{1}{n}\sum_{i=1}^{n} X_i$ приближается к нормальному закону.***
- Таким образом, имеем следующую асимптотическую формулу, являющуюся математическим выражением теоремы Ляпунова:

$$P\left\{\left|\frac{1}{n}\sum_{i=1}^{n} X_i - \frac{1}{n}\sum_{i=1}^{n} M(X_i)\right| < t\sqrt{\frac{\overline{D(X)}}{n}}\right\} = \frac{2}{\sqrt{2\pi}}\int_0^t e^{-\frac{x^2}{2}}\,dx = \Phi(t),$$

где $\overline{D(X)} = \frac{1}{n}\sum_{i=1}^{n} D(X_i)$ - осредненная дисперсия.

При числе испытаний n, стремящемуся к бесконечности, среднее арифметическое сходится по вероятности к математическому ожиданию.

- **Пример.** Дисперсия каждой из 400 независимых случайных величин равна 25.
- Найти вероятность того, что абсолютное отклонение средней арифметической случайных величин от средней арифметической их математических ожиданий не превысит 0,5.
- ***Решение.***
- По условию задачи $n = 400$, $D(X_i) = 25$ и $\varepsilon = 0{,}5$.
- Определим осредненную дисперсию:

$$\overline{D(X)} = \frac{1}{n}\sum_{i=1}^{n} D(X_i) = \frac{1}{n} n \cdot D(X_i) = D(X_i) = 25.$$

- Исходя из значения $\varepsilon = 0{,}5$ определим t:

$$t = \frac{\varepsilon}{\sqrt{\frac{\overline{D(X)}}{n}}} = \frac{0{,}5}{\sqrt{\frac{25}{400}}} = \frac{0{,}5 \cdot 20}{5} = 2.$$

- Искомая вероятность по теореме Ляпунова равна

$$P\left\{\left|\frac{1}{n}\sum_{i=1}^{n} X_i - \frac{1}{n}\sum_{i=1}^{n} M(X_i)\right| < 0{,}5\right\} = \Phi(2) = 0{,}9545.$$