**1. Что называется выборкой? Что можно сказать об элементах выборки?**

**Выборка** — это конечное множество наблюдений, полученных из **генеральной совокупности**, то есть из всей совокупности возможных значений случайной величины.  
Цель выборки — сделать выводы о свойствах этой совокупности, не исследуя её полностью.

**Элементы выборки** считаются:

* **Случайными** — они подчиняются случайным законам, зависят от вероятности.
* **Независимыми** — значение одного элемента не влияет на другие.
* **Одинаково распределёнными** — все они подчиняются одному и тому же закону распределения.

Эти три свойства (i.i.d. — independently and identically distributed) важны для корректности большинства статистических методов.

**2. Что такое статистика?**

**Статистика** (в широком смысле) — это наука, изучающая способы **сбора, обработки, анализа и интерпретации данных** о массовых явлениях, происходящих в природе, обществе и технике.

**Статистика** (в узком смысле) — это числовая характеристика, вычисляемая по данным выборки.  
Это может быть:

* **Выборочное среднее** — аналог математического ожидания, отражает «средний» уровень данных.
* **Выборочная дисперсия** — мера разброса значений вокруг среднего.
* **Медиана** — значение, делящее выборку пополам.
* **Мода** — наиболее часто встречающееся значение.

Важно: статистика зависит **только от данных**, но **не зависит от неизвестных параметров** самого распределения. Она используется для построения оценок этих параметров.

**3. Почему результаты полученные по элементам выборки можно считать близкими к истинным значениям распределения?**

Потому что действует **закон больших чисел**: при увеличении числа наблюдений (размера выборки) выборочные характеристики **сходятся к теоретическим (истинным) параметрам**.

Например:

* Среднее значение по выборке приближается к математическому ожиданию.
* Выборочная дисперсия приближается к истинной дисперсии распределения.

Это означает, что достаточно большая выборка даёт статистику, близкую к «истине» — даже если мы не знаем сам закон распределения.

**4. Что такое вариационный ряд?**

**Вариационный ряд** — это **упорядоченный** (обычно по возрастанию) список всех наблюдаемых значений в выборке. Если значение повторяется, оно включается столько раз, сколько встретилось.

Пример: для выборки [3, 1, 2, 2], вариационный ряд будет: [1, 2, 2, 3].

Иногда рядом указывают **частоту** (сколько раз встречается каждое значение) или **относительную частоту** (доля от общего числа).

Вариационный ряд позволяет:

* наглядно видеть распределение данных,
* находить медиану, моду, размах (макс − мин),
* строить диаграммы (гистограммы, полигоны и др.).

**5. Как строится группированная выборка?**

**Группированная выборка** — это представление данных, где значения не перечисляются поштучно, а **разбиваются на интервалы**, и для каждого интервала считается, сколько значений в него попало (**частота**). Это особенно полезно при большом объеме данных.

Пошагово:

1. **Находим минимум и максимум выборки** — определяем диапазон значений.
2. **Определяем число интервалов** kkk, обычно по формуле **Стерджеса**: k = ln(n) + 1

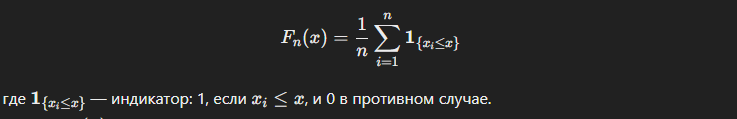
где n — размер выборки.

1. **Делим диапазон на k равных интервалов** — рассчитываем ширину.
2. **Строим интервалы** вида: [ ; )
3. **Считаем частоту** попадания значений выборки в каждый интервал.

Результат — таблица с интервалами и соответствующими частотами, которая показывает **структуру распределения данных в обобщенном виде**.

**6. Как получить оценку функции распределения СВ (эмпирическую функцию распределения)?**

**Эмпирическая функция распределения (ЭФР)** строится по выборке и показывает долю значений, не превышающих заданное число x. Формально:

  
То есть Fn(x) — это **доля наблюдений**, меньших или равных x. ЭФР всегда неубывающая, ступенчатая функция. Она приближает истинную функцию распределения.

**7. Что такое полигон?**

**Полигон частот** — это **ломаная линия**, соединяющая середины оснований соседних столбцов гистограммы, по оси абсцисс откладываются середины интервалов, а по оси ординат — частоты или относительные частоты.

Если использовать **относительные частоты**, получается **полигон относительных частот**. Он служит наглядной оценкой формы распределения и является более плавным аналогом гистограммы.

**8. Оценкой какой функции является гистограмма?**

**Гистограмма** является **оценкой функции плотности вероятности** непрерывной случайной величины.  
Она показывает, как часто значения выборки попадают в заданные интервалы, и при нормировке (делении на ширину интервалов и число наблюдений) приближает **плотность распределения**.

**9. В чем состоит задача оценивания?**

**Задача оценивания** состоит в том, чтобы по результатам выборки получить приближённое значение **неизвестного параметра** генерального распределения.  
Цель — построить **оценку параметра** (например, математического ожидания, дисперсии, вероятности и т.д.) с помощью статистик, вычисленных из данных.

Оценка должна быть:

* **Несмещённой** (в среднем правильной),
* **Состоятельной** (приближается к истинному значению при n стремящемся к беск.),
* **Эффективной** (имеет минимальную возможную дисперсию).

**10. Что такое оценка параметра?**

**Оценка параметра** — это числовое значение, вычисленное по выборке, которое приближённо характеризует неизвестный параметр распределения (например, среднее, дисперсию, вероятность успеха и т.д.).

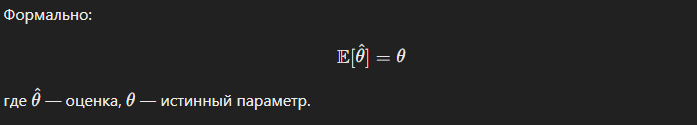
Бывает:

* **Точечная оценка** — одно число, например: μ^=5.2,
* **Интервальная оценка** — диапазон, в котором с заданной вероятностью находится параметр, например: μ∈[4.9;5.5] при доверии 95%.

Оценка параметра должна быть построена из **статистики**, то есть зависеть только от выборки.

**11. Дать определение несмещенной оценки**

**Несмещённой оценкой** параметра называется такая статистика, **математическое ожидание которой равно истинному значению параметра**.



**12. Дать определение асимптотически несмещенной оценки**

**Асимптотически несмещённая оценка** — это такая оценка, у которой **смещение стремится к нулю при увеличении объема выборки.**



То есть она приближается к истинному значению при n -> inf, даже если при конечном n может быть смещённой.

**13. Дать определение состоятельной оценки**

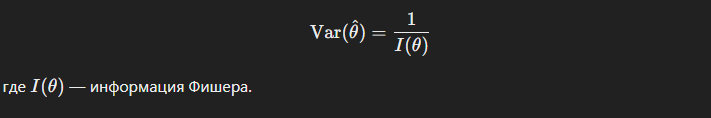
**Состоятельной** называется оценка, которая **с вероятностью 1** стремится к истинному значению параметра при n→inf:



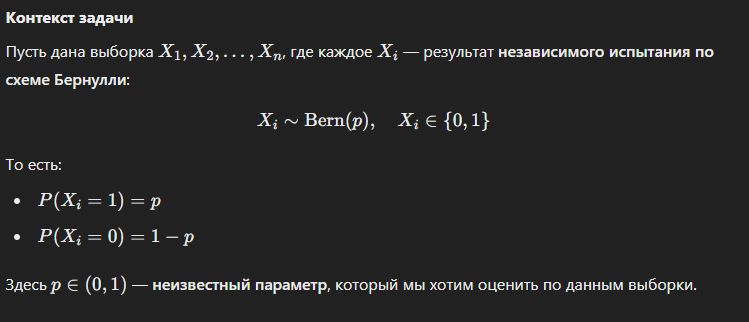
Это означает, что с ростом выборки оценка «почти наверняка» становится всё ближе к параметру

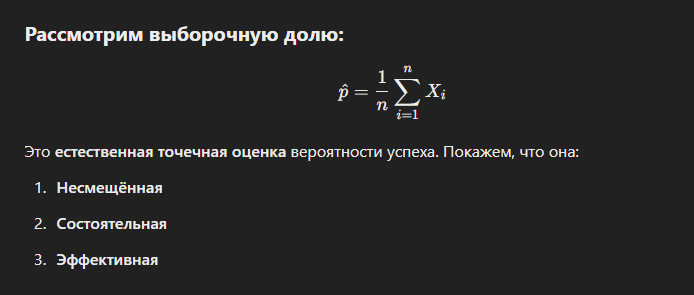
**14. Дать определение эффективной оценки**

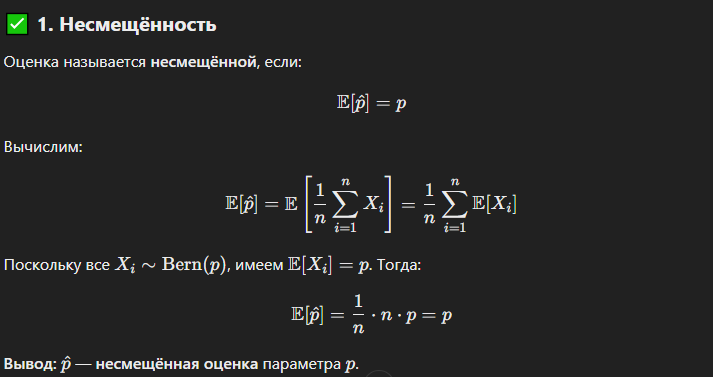
**Эффективная оценка** — это несмещённая оценка, имеющая **наименьшую возможную дисперсию** среди всех несмещённых оценок параметра.  
Если оценка достигает **границы Крамера–Рао**, она считается эффективной:

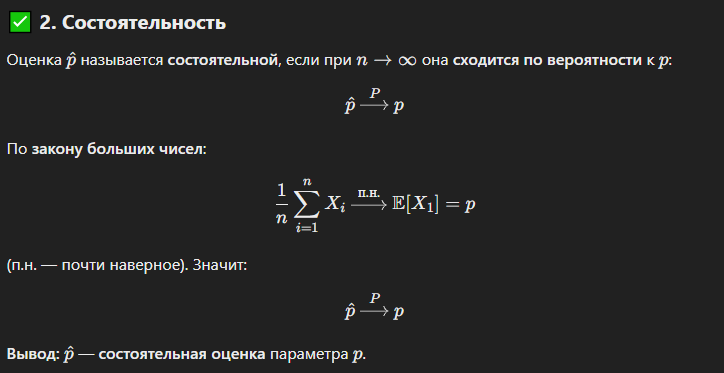


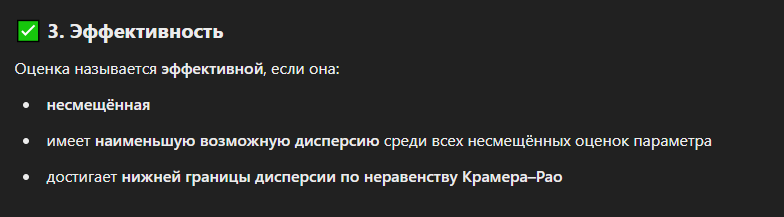
**15. Что является несмещенной, эффективной и состоятельной оценкой вероятности (доказать)?**

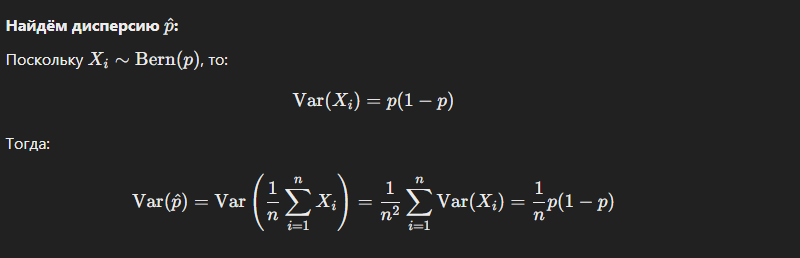






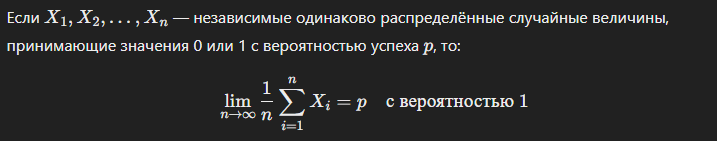




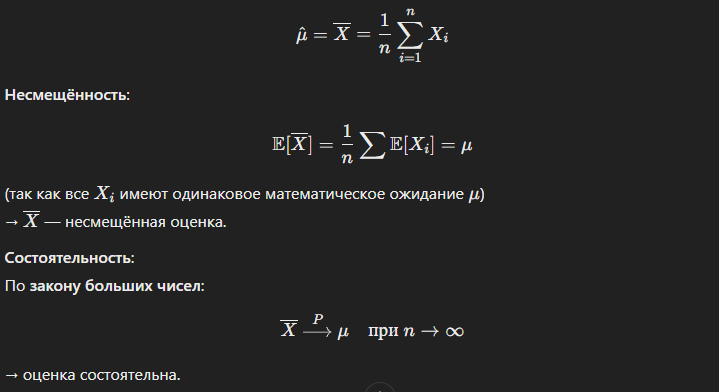


**16. Закон больших чисел (теорема Бернулли)**

Относительная частота успехов в серии испытаний сходится к вероятности успеха p по мере роста числа испытаний.

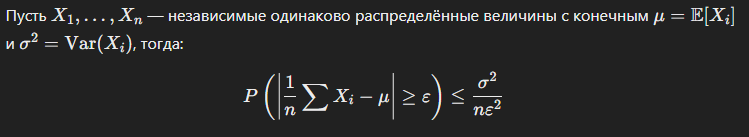


**17. Оценивание математического ожидания (обосновать несмещенность и состоятельность)**

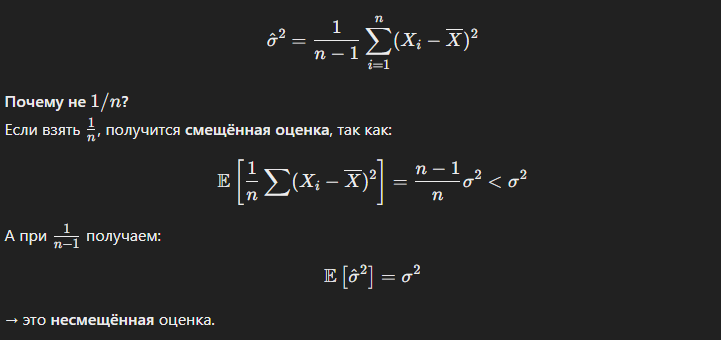


**18. Закон больших чисел (теорема Чебышева)**

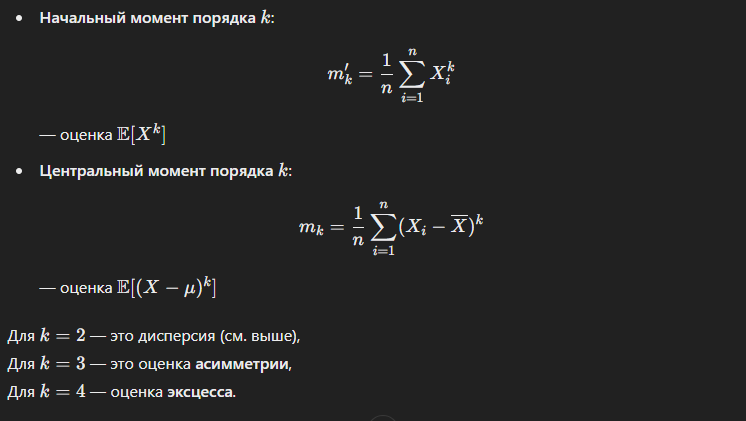
Вероятность того, что среднее сильно отклоняется от ожидания, убывает при росте nnn. Это обоснование сходимости средней к ожиданию.



**19. Оценивание дисперсии. Несмещенная оценка дисперсии с выводом**

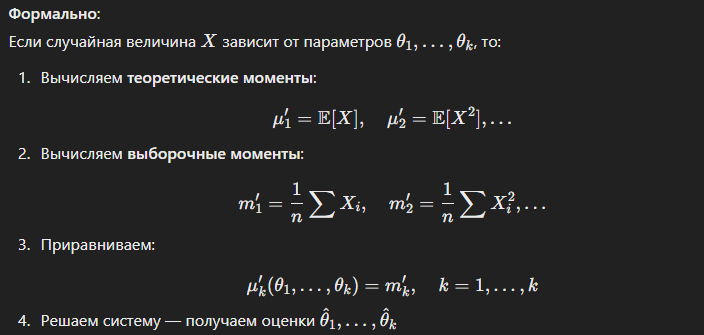


**20. Оценивание начальных и центральных моментов СВ по выборке**



**21. Суть метода моментов и его формальное описание**

Метод моментов заключается в том, чтобы **приравнять теоретические моменты** (выраженные через параметры) к **выборочным** и решить полученную систему уравнений для оценки параметров распределения.

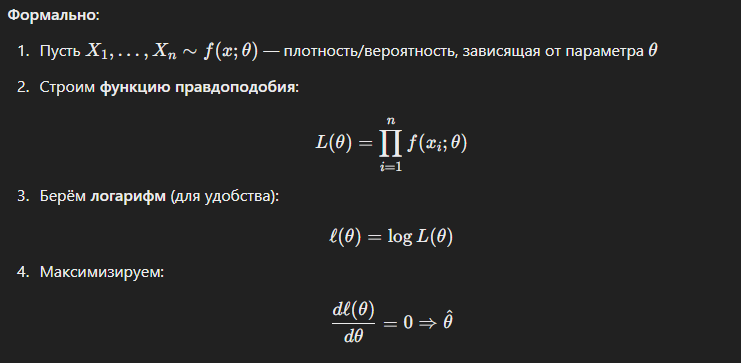


**22. Пример оценки параметров по методу моментов для ДСВ и для НСВ (можно из своей РГР)**

Дискретная случайная величина – смотреть п. 3. Непрерывная случайная величина – смотреть п.5.

**23. Суть метода максимального правдоподобия и его формальное описание**

Метод максимального правдоподобия заключается в выборе таких значений параметров, **при которых вероятность наблюдать именно эту выборку максимальна**.



**24. Пример оценки параметров по методу максимального правдоподобия для ДСВ и для НСВ (можно из своей РГР)**

Дискретная случайная величина – смотреть п. 3. Непрерывная случайная величина – смотреть п.6.