19 октября. Вводное занятие «Верю – не верю». Правдоподобие. Зачем нужна теория вероятностей. Закон больших чисел

1. Перчатки в ящике

Хорошо известна задача: в ящике 100 левых и 100 правых перчаток. Сколько нужно вынуть перчаток (наугад), чтобы наверняка получилась пара?

Ответ прост и получается из принципа Дирихле: нужно достать 101 перчатку.

Если заменить требование «наверняка» требованием «почти наверняка», то оказывается, что достаточно взять всего 6 перчаток и тогда с очень высокой вероятностью среди них будет пара.

Доказательство. Предположим, что первая вынутая перчатка – левая. Какова теперь вероятность того, что оставшиеся пять тоже будут левыми?

$$\frac{99}{199}\cdot\frac{98}{198}\cdot\frac{97}{197}\cdot\frac{96}{196}<\left(\frac{1}{2}\right)^5=\frac{1}{32}=0,03125.$$

Это значит, что с вероятностью больше, чем $1-0,03125=0,96875$ не все перчатки будут левыми, то есть среди них найдется пара.

Это нас устроит, *если мы верим в событие*, имеющее вероятность $0,96$ или больше. Считаем его очень *правдоподобным*. А событие с вероятностью около 0,03 считаем *неправдоподобным* и не принимаем его в расчет.

Если кто-то так не считает, пусть возьмет не 6 перчаток, а 7.

2. В чем разница между «наверняка» и «практически наверняка»

В задаче про перчатки мы получили пару, взяв 6 перчаток. Но не наверняка, а практически наверняка. *В жизни между «наверняка» и «практически наверняка» разницы нет*. В однократном опыте мы верим в очень вероятные события и не верим в практически невероятные.[1]

Важно уметь выбрать эту «решающую» вероятность. В примере с перчатками, например, можно было выбрать 0,95 или даже 0,96. Самая большая неприятность, что может случиться – не будет пары перчаток.

Но если (в медицине или в авиации) цена ошибки очень велика, то лучше принимать на веру только те события, вероятности которых намного выше, скажем не менее 0,999. Выбор только за нами.

[1] Обратите внимание: речь идет об однократном опыте. Если такой опыт проводить множество раз, то рано или поздно случится и самое невероятное. Пословица гласит: «Раз в жизни стреляет незаряженное ружье».

Вот примеры задач, которые нельзя решить наверняка, но можно решить практически наверняка:

1. Сколько нужно запасти саженцев, чтобы не менее 100 из них прижилось, если известна вероятность гибели одного саженца: 0,05.
2. Сколько нужно иметь в запасе одеял в городском убежище на случай землетрясения или наводнения?
3. Сколько нужно иметь в запасе бумаги, чтобы хватило всем участникам ЕГЭ по математике?

Ни одна из этих задач не может быть решена точно, поскольку число погибших саженцев может быть любым, школьники могут исписать немыслимую гору бумаги и т.п. Но все эти задачи разумно решаются с помощью теории вероятностей.

Теория вероятностей важна, поскольку она может *предложить полезные решения задач* «почти наверняка» там, где «точная математика» решений не дает или дает бессмысленные (например, 101 перчатку).

3. Fish or chicken

Пример вероятностной задачи. В самолете 200 пассажиров, и авиакомпания загрузила на борт 100 порций обедов с курицей и 100 порций – с рыбой. Не все любят курицу, не все любят рыбу. Может случиться (и иногда бывает), что некоторые пассажиры недовольны.

Авиакомпания провела исследование и выяснила, что на 100 последних одинаковых рейсах всего оказалось 738 недовольных пассажиров. Значит, среднее число недовольных в расчете на один рейс, равно 7,38.

Можно ли опираться на это среднее значение при дальнейших прогнозах?

Мнения участников кружка разделились. После обсуждения мы пришли к следующим утверждениям:

1. Среднее число недовольных пассажиров 7,38 – это *статистика*, которая в дальнейшем может изменяться (на других 100 таких же рейсах среднее число недовольных может быть другим). То есть, среднее число – величина случайная и изменчивая.
2. Мы не верим в то, что изменения будут значительны. Мы готовы поверить в 5, 6 или 10 недовольных в среднем пассажиров, но не готовы поверить в 20 или 50. Нам представляется *неправдоподобным* такое резкое изменение среднего значения. Почему? Мы это чувствуем.

3. Если среднее все же сильно меняется, это значит, что изменились условия эксперимента (например, изменилось число пассажиров на рейсе или вдруг среди них стало много вегетарианцев).

Иными словами, если условия эксперимента (полета) остаются неизменными, то при накоплении большого числа наблюдений над величиной, можно рассчитывать на то, что среднее значение этой величины в дальнейшем значительно меняться не будет – наблюдается *статистическая устойчивость*. Почему? Только наш опыт и интуиция утверждают, что это так.

4. Проверка устойчивости с помощью монеты

При бросании монеты решка и орел имеют равные шансы на выпадение. По этой причине мы думаем, что из 10 брошенных монет, *в среднем* 5 должно выпасть орлом вверх. Мы почти не верим в 0 орлов или в 10 орлов, слабо верим в 1 или в 9 орлов и даже в 2 или в 8 орлов, но вполне допускаем 3, 4, 5, 6 или 7 орлов.

Каждый из участников кружка бросил монету 200 раз. По одной из таких серий был построен график. Он получился примерно таким:

Орел выпал 104 раза из 200. Частота орла 0,52.

Если повторять такой эксперимент много раз, то можно увидеть общую закономерность: колебания частоты вначале очень большие, а затем наступает стабилизация – колебания все меньше и меньше, *частота стабилизуется около вероятности* 0,5.

К концу эксперимента частота выпадения орла, как правило, отличается от 0,5, но большие отличия встречаются реже, чем малые. Например, частота может оказаться 0,55 или 0,43, но это гораздо реже, чем 0,51 или 0,49.

1. Почему колебания частоты постепенно затухают?
2. Почему частота стабилизируется около вероятности 0,5?
3. Почему вероятности больших отличий частоты от вероятности малы?
4. Можно ли считать правдоподобным, что частота отклонится от вероятности на 0,01? А на 0,1? А на 0,2?

Утверждения, которые содержатся в этих вопросах, представляют собой важнейший закон статистики и вероятности: *закон больших чисел*. Он проявляется по-разному. Например, так.

При росте числа одинаковых экспериментов, *частота события* испытывает все меньшие и меньшие колебания и *постепенно стабилизируется вблизи вероятности этого события*.

Аналогично:

При росте числа одинаковых экспериментов *среднее арифметическое* наблюдений испытывает все меньшие колебания и постепенно *стабилизируется вблизи математического ожидания* соответствующей случайной величины.

Закон больших чисел дает нам в руки очень мощное оружие:

1. *Зная вероятность события, можно предсказывать его частоту* и даже правдоподобные отклонения.
2. *Зная частоту события, можно оценивать его вероятность*, опять-таки с правдоподобной ошибкой.

В этом занятии участники столкнулись с большим числом незнакомых слов и понятий: частота, отклонение, математическое ожидание и т.п.

Во-первых, есть консультация. Во-вторых, есть учебники. А в-третьих, для того и есть кружок, чтобы постепенно во всем разобраться.

В качестве домашнего задания участникам было предложено провести с помощью модуля coins.exe (странице «Эл.ресурсы/компьютерные модели»: http://ptlab.mccme.ru/node/187) серии бросаний монет и, наблюдая получающиеся графики частот, выделить и описать своими словами общие закономерности.

И.Р.Высоцкий