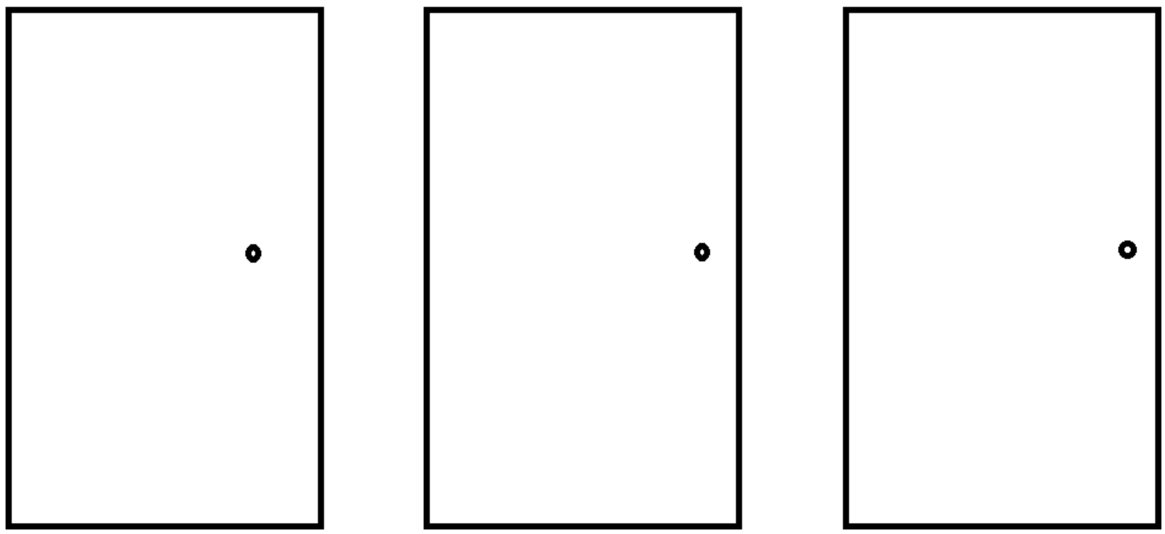
KARPOV.COURSES >>>
КОНСПЕКТ

> Конспект > 3 урок > ТЕОРИЯ ВЕРОЯТНОСТЕЙ

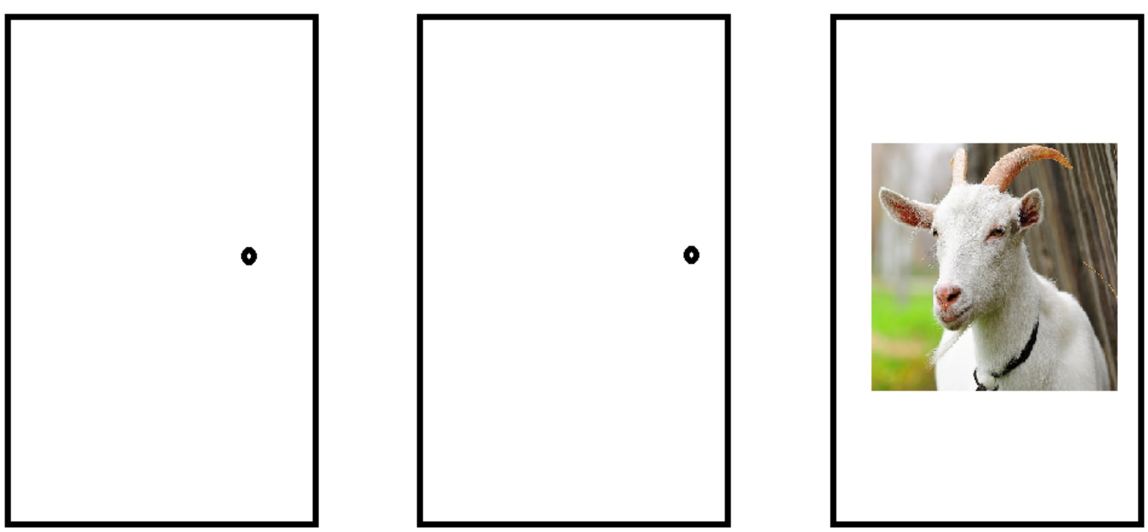
> Оглавление

> Парадокс Монти-Холла

Оригинальная ситуация, в честь которой назвали парадокс, выглядит так.Представьте, что вы участвуете в шоу наподобие Поля Чудес и вам предложили получить автомобиль! Приз находится за одной из трёх дверей - за двумя другими находятся козы. Где что - вы не знаете.

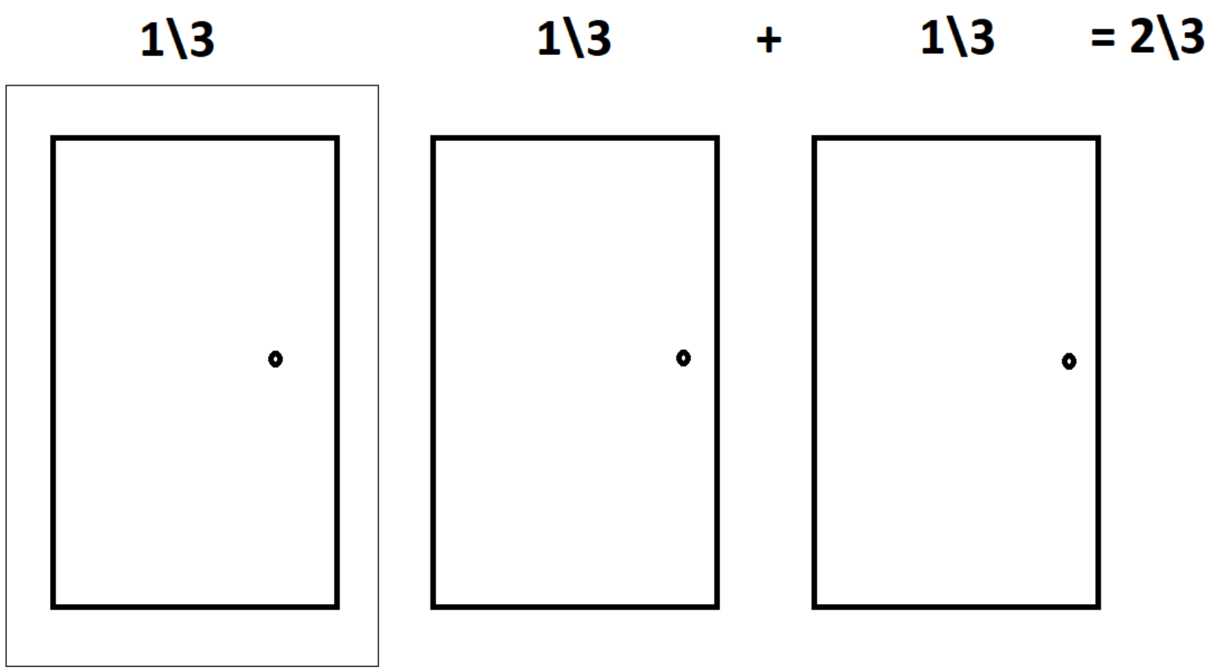


Допустим, что вам действительно нужен автомобиль и не нужна коза (всякое бывает). Вы рандомно выбираете одну из дверей и ждёте ответа ведущего. Но ведущий вместо ответа открывает одну из невыбранных дверей, а там стоит коза.

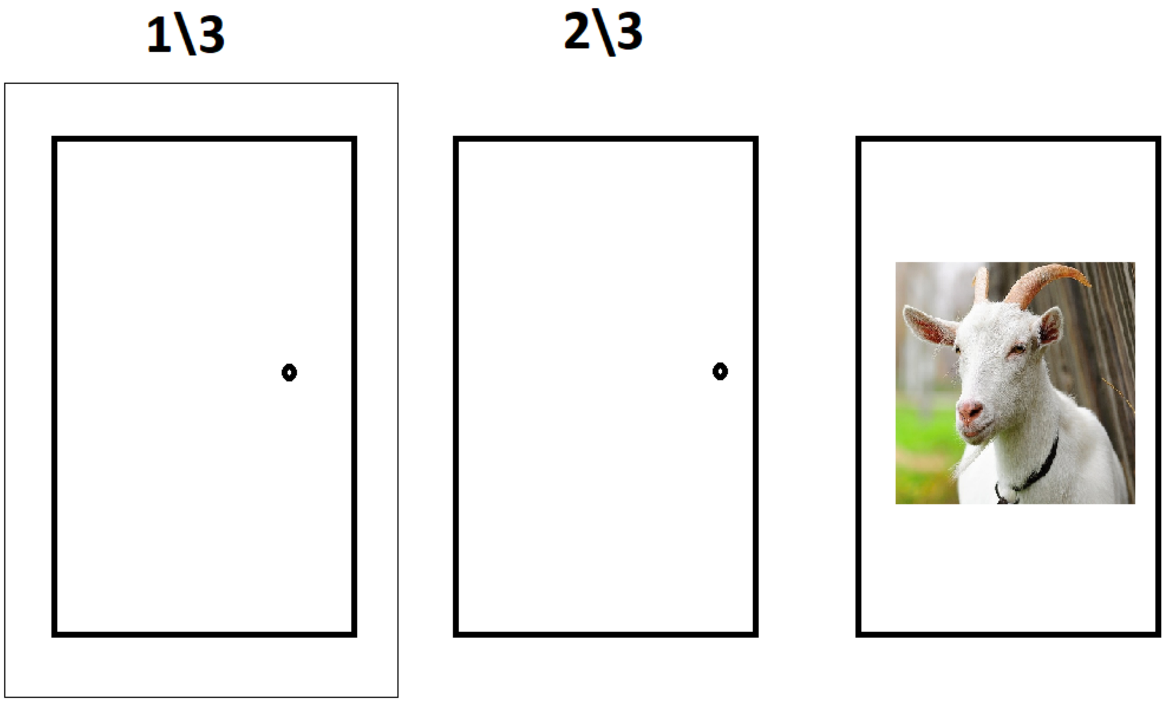


Ведущий предлагает вам выбор. Раз вы теперь точно знаете, за какой дверью стоит коза, поменяете ли вы дверь? Многие остаются с той дверью, которую они выбрали изначально.

Парадокс заключается в том, что всегда выгоднее изменить свой выбор! Как это получается? Давайте посмотрим на вероятность того, что машина находится не за той дверью, которую мы изначально выбрали:



А теперь фокус: от того, что третья дверь "выбыла из игры", распределение вероятностей никак не поменялось! Это всё те же $\frac{1}{3}$ и $\frac{2}{3}$- просто теперь последняя вероятность целиком и полностью принадлежит одной невыбранной двери:



Соответственно, выбирать её всегда выгоднее. Тем не менее, не умудрённые опытом люди часто ошибаются на этом моменте. В этом коварство **условных вероятностей** - условных не потому, что они не существуют. Они отражают вероятность события **при условии, что произошло другое событие**. Ситуаций, требующих условных вероятностей, в реальной жизни очень много, и сейчас мы научимся считать их формально.

Ещё немного о Монти-Холле

Анатолий в лекции предлагает вам самостоятельно запрограммировать эту ситуацию. Если вам лень придумывать код, то вот его пример. Попробуйте расширить эту симуляцию для большего количества дверей :)

Как вы могли понять, в этом парадоксе есть очень большой психологический компонент. Почитать научную обзорную статью о возможных механизмах ошибки при ситуации Монти-Холла можно вот тут.

> Формула Байеса

Названа в честь священника, который её и придумал. Выглядит она так:

$$P(Y|X) = \frac{P(Y)P(X|Y)}{P(X)}$$

$P(Y)$ и $P(X)$ - это вероятности событий Y и X, соответственно.

$P(Y|X)$- вероятность события Y, **если наступило событие X**

$P(X|Y)$ - вероятность события X, **если наступило событие Y**

Очень важно не путать $P(Y|X)$ и $P(X|Y)$! Они могут казаться похожими, но очень редко совпадают в реальной жизни. Скажем, если с вами рядом летает комар, то довольно высока вероятность получить укус - **P(получить укус|рядом комар)**. Но если вы получили укус, то это необязательно сделал именно комар - вас могли укусить любая другая двукрылая пакость, собака, змея или излишне любвеобильный партнёр - **P(рядом комар|получить укус)**.

Пример из лекции

У нас есть две группы стрелок из лука - одна из профессионалов (Y_1), другая из новичков (Y_2). Допустим, профессионалов 5, и их точность попадания - 8 из 10 мишеней. Новичков же 15, и их точность - 4 из 10 мишеней.

Мы подошли к мишени и увидели, что в неё попали. Какова вероятность, что это попадание сделал профессионал?

Для начала подсчитаем вероятность того, что случайно выбранный стрелок принадлежит к группе профессионалов:

$$P(Y_1) = \frac{5}{20} = 0.25$$

Вероятность попадания профессионала в мишень нам уже известна:

$$P(X|Y_1) = 0.8$$

Теперь рассчитаем вероятность попадания в мишень в целом:

$$P(X) = P(Y_1)P(X|Y_1) + P(Y_2)P(X|Y_2) = 0.25 * 0.8 + 0.75 * 0.4 = 0.5$$

Обратите внимание, что формула выше легко приводится к уже знакомой нам формуле $P = \frac{n}{m}$:

$$\frac{P(Y_1)P(X|Y_1)}{P(Y_1)P(X|Y_1) + P(Y_2)P(X|Y_2)}$$

В итоге у нас получается такая вероятность:

$$P(Y|X) = \frac{P(Y)P(X|Y)}{P(X)} = \frac{0.25 * 0.8}{0.5} = 0.4$$

Что ещё почитать:

Для дополнительного погружения в идею формулы Байеса можно почитать эту заметку.

Также Анатолий много времени посвящает медицинскому примеру с парадоксом Байеса - один из вариантов истории с ним можно прочитать вот тут.

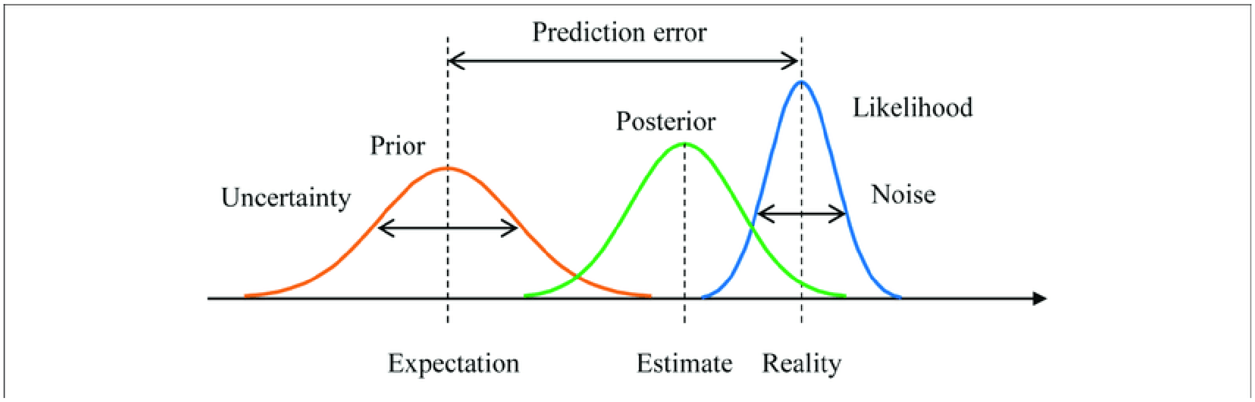
> Байесовская статистика

Помните, мы в первом конспекте упоминали **субъективную интерпретацию** вероятности? Она отражает степень убеждённости в конкретном событии или факте жизни. И формальная реализация этого типа вероятности очень многим обязана именно теореме Байеса! В частности, из неё родилось **байесианство** как течение философии науки и **байесовская статистика** как один из двух доминирующих подходов к статистическому выводу.

Вся она базируется на трёх идеях - априорной и апостериорной вероятностях, а также функции правдоподобия. Относительно теоремы Байеса они соотносятся как $P(X)$, $P(Y|X)$и $P(X|Y)$ соответственно. Объясняя простым языком, принцип такой:

1. Мы задаём наши предположения о распределении какого-то параметра (априорная вероятность)
2. Мы получаем реальное распределение на основе полученных в исследовании данных (функция правдоподобия)

3. Комбинируя наши изначальные убеждения и эмпирически полученный результат, мы приходим к новому убеждению (апостериорная вероятность)



В этом курсе мы не будем затрагивать байесовские методы и будем учить вас классической, частотной парадигме статистики. Но если вам интересно, как этим всем пользоваться и почему это может быть круто, то вот подборка статей и книжек:

- Размышления на тему того, почему байесовский подход может быть лучше, с рядом других ссылок
- Введение в байесовские методы на Хабре
- Открытые материалы курса по байесовским методам машинного обучения
- Книжка про байесовскую статистику на Python
- Другая очень хорошая и подробная книжка, но на R
- Немного про A/B тестирование в обеих парадигмах