The **PRIOR** hypothesis:
The original probability of the hypothesis without any additional information

The **LIKELIHOOD** interpreted as:
P(observation GIVEN the hypothesis)

$$P[A \mid B] = \frac{P[A]P[B \mid A]}{P[B]}$$

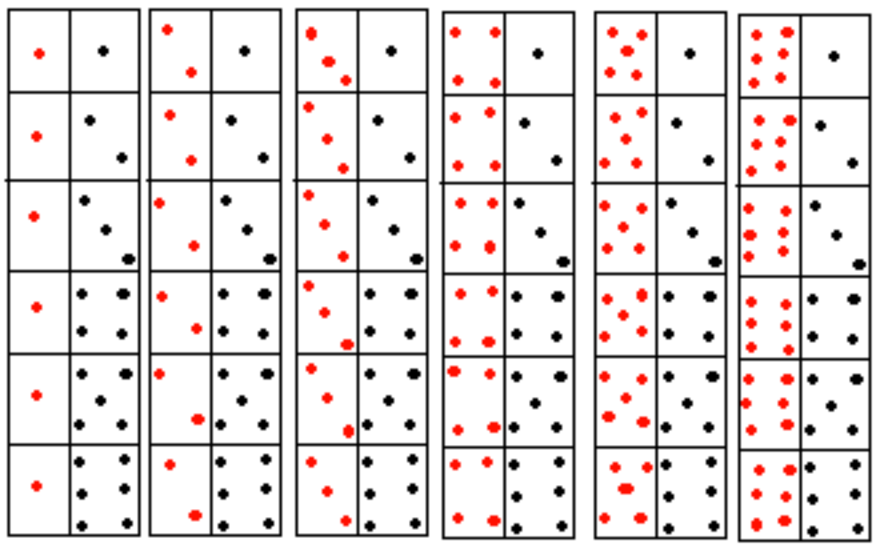
The **observation/data/ Evidence** that has been observed

the **POSTERIOR probability** sometimes interpreted as the P(hypothesis GIVEN the observation)

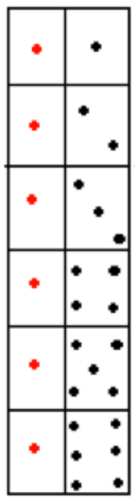
Venn diagram of Bayes'

<u>probability that two dice will sum to three?</u>

- *You have no new information*
- *You have to consider the entire universe*
- *P[Sum to three] =2/36*



- You have updated information
- Red die=1
- P[Sum to three|red=1]=1/6



Bayes

Example: Suppose we want to calculate the probability that someone will die of lung cancer given that they smoke. We study a cohort of individuals, determining which ones smoke and which ones do not and track them until they died. Then we could calculate the number of smokers who died of lung cancer.

There is an easier way, however….

USE BAYES

Example: Suppose we want to calculate the probability that someone will die of lung cancer given that they smoke. We study a cohort of individuals, determining which ones smoke and which ones do not and track them until they died. Then we could calculate the number of smokers who died of lung cancer.

There is an easier way, however….

USE BAYES!

$$\boxed{P[A \mid B] = \frac{P[A]P[B \mid A]}{P[B]}}$$

Example: Suppose we want to calculate the probability that someone will die of lung cancer given that they smoke.

Answer:

Specify the question: What is event 'A' and what is event 'B'?

$$\boxed{P[A \mid B] = \frac{P[A]P[B \mid A]}{P[B]}}$$

Example: Suppose we want to calculate the probability that someone will die of lung cancer given that they smoke.

Answer:

Specify the question: What is event 'A' and what is event 'B'?

P[Death due to lung cancer|smoker]

This means that Bayes formula will be:

$$\boxed{P[Lung\_Cancer\_Death \mid Smoker] = \frac{P[Lung\_Cancer\_Death]\,P[Smoker \mid Lung\_Cancer\_Death]}{P[Smoker]}}$$

<u>Example:</u> Suppose we want to calculate the probability that someone will die of lung cancer given that they smoke.

<u>Answer:</u>

Specify the question: What is event ‘A’ and what is event ‘B’?

P[Death due to lung cancer|smoker]

This means that Bayes formula will be:

$$\boxed{P[Lung\_Cancer\_Death \mid Smoker] = \frac{P[Lung\_Cancer\_Death]\,P[Smoker \mid Lung\_Cancer\_Death]}{P[Smoker]}}$$

The probabilities on the right are already present in public databases. If we use Bayes, there is no need for costly study.

Example: Suppose we want to calculate the probability that someone will die of lung cancer given that they smoke.

Answer:

Specify the question: What is event ‘A’ and what is event ‘B’?

P[Death due to lung cancer|smoker]

Estimated from death records

$$\boxed{P[Lung\_Cancer\_Death|Smoker] = \frac{P[Lung\_Cancer\_Death]\,P[Smoker|Lung\_Cancer\_Death]}{P[Smoker]}}$$

Polling appropriate control population

Example: Suppose we want to calculate the probability that someone will die of lung cancer given that they smoke.

Answer:

Specify the question: What is event 'A' and what is event 'B'?

P[Death due to lung cancer|smoker]

P[smoker] =0.5
P[smoker|Death due to lung cancer] = 0.9
P[Death due to lung cancer] = 0.3

$$P[\text{Death due to lung cancer}|\text{smoker}] = \frac{0.9 \times 0.3}{0.5} = 0.54$$

This study also gives the
P[Death due to lung cancer|Non-smoker] = 0.06

Bayesian Probability:

- Tension between frequentists and bayesians
  → Where does the prior come from?
    * all data that is external to the current study
    * mitigate this issue: use non-informative prior

- Most events are not repeatable so bayesian probability attempts to utilize a personal assessment of an outcomes likelihood
    - Do this by weighting probability with a prior probability
        - Some scientists may use different priors!
        - Increasing data/information means that the prior should have less impact on the posterior

    - This may seem subtle but it is quite different from what science traditionally does:

    The parameter is treated as though it were a random variable instead of a constant truth/value

<u>Bayesian Probability:</u>

Bayes Factor:

- *Allows strength of two competing hypotheses (a null and an alternate) to be quantified*

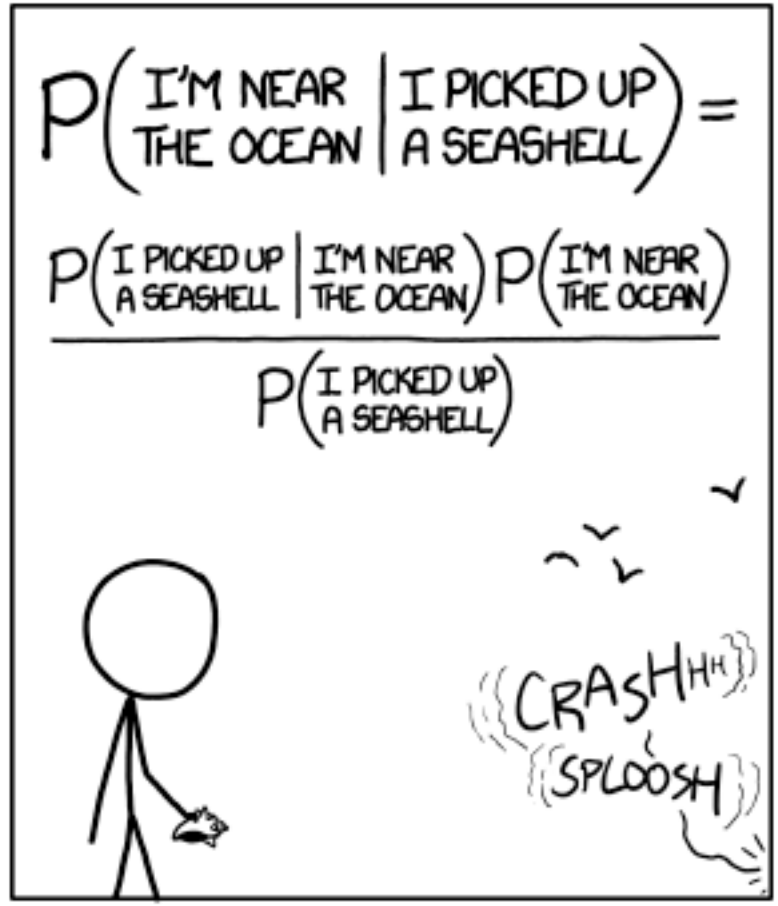
- *Both hypotheses must use same prior*

- *Null hypothesis isn't assumed (it doesn't have primacy over the alternative hypothesis)*

- *Similar to likelihood that we will see towards the end of this course*

Bayes

http://xkcd.com/1236/



STATISTICALLY SPEAKING, IF YOU PICK UP A SEASHELL AND *DON'T* HOLD IT TO YOUR EAR, YOU CAN PROBABLY HEAR THE OCEAN.