PROBABILITY AND STATISTICS: A RAPID RECAP

CS1951A INTRO TO DATA SCIENCE

TIM KRASKA <TIM_KRASKA@BROWN.EDU>

teaching datascience .org

OUTLINE

1. **Probability Spaces & Probability Functions**
   1. Example: Rolling a Die
   2. Conditional Probability
   3. Independent Events
2. **Bayesian Statistics**

STATISTICS ≠ PROBABILITY THEORY

Probability theory: mathematical theory that describes uncertainty.

Statistics: set of techniques for extracting useful information from data.

PROBABILITY SPACE

A probability space has three components:

(1) A sample space Ω which is the set of all possible outcomes of the random process modeled by the probability space;
(2) A family of sets F representing the allowable events, where each set in F is a subset of the sample space Ω;
(3) A probability function $\Pr : F \rightarrow R$ satisfying the definition below:

An element of Ω is a simple event. In a discrete probability space, we use $F = 2^{\Omega}$

PROBABILITY FUNCTION

A probability function is any $\Pr : F \to R$ that satisfies the following conditions:

① For any event E, $0 \le Pr(E) \le 1$;

② $Pr(\Omega) = 1$;

③ For any finite or countably infinite sequence of pairwise mutually disjoint events $E_1, E_2, E_3 \ldots$

$$\boxed{\Pr\left(\bigcup_{i \le 1} E_i\right) = \sum_{i \ge 1} P\mathrm{r}(E_i)}$$

The probability of an event is the sum of the probabilities of its simple events.

EXAMPLE: TOSSING A (FAIR) COIN

$$\mathrm{W} = \{H, T\}$$

$$F = 2^{\mathrm{W}} = 2^2 = 4\,Events$$

$$F = \{\{\ \}, \{H\}, \{T\}, \{H, T\}\}$$

$$\Pr(\{\ \}) = 0$$

$$\Pr(\{H\}) = 0.5$$

$$\Pr(\{T\}) = 0.5$$

$$\Pr(\{H, T\}) = 1$$

EXAMPLE: ROLLING A DIE

$\Omega = \{1,2,3,4,5,6\}$

$F = 2^{\Omega} = 2^{6}\ Events$

$\Pr(\{\ \}) = 0$

$\Pr(\{1\}) = \Pr(\{2\}) = \Pr(\{3\}) = \Pr(\{4\}) = \Pr(\{5\}) = \Pr(\{6\}) = \frac{1}{6}$

$\Pr(\{1,2\}) = \Pr(\{1,3\}) = \Pr(\{1,4\}) = \Pr(\{1,5\}) = \Pr(\{1,6\}) = \frac{2}{6}$

...

COMPUTING CONDITIONAL PROBABILITY

The **conditional probability** that event E1 occurs given that event E2 occurs is:

$$\boxed{\Pr(E_1 \mid E_2) = \frac{\Pr(E_1 \cap E_2)}{\Pr(E_2)}}$$

The conditional probability is only well-defined if $\Pr(E_2) > 0$

By conditioning on E2 we restrict the sample space to set E2.

Thus we are interested in $\Pr(E_1 \cap E_2)$ normalized by $\Pr(E_2)$.

EXAMPLE: CONDITIONAL PROBABILITY

We have two coins: A is a fair coin, B has probability 2/3 to come up as HEAD. We chose a coin at random and got HEAD.

What is the probability that we chose coin A?

① E_1 = the event "chose coin A"
② E_2 = the event "outcome is HEAD"

Conditional probability that we chose coin A **given** that the outcome is HEAD is denoted: $Pr(E_1 \mid E_2)$

EXAMPLE: CONDITIONAL PROBABILITY

Define a **sample space** of **ordered** pairs: *(coin, outcome)*

The same space has four points:

① $\{(A,h), (A,t), (B,h), (B,t)\}$

② $Pr((A,h)) = Pr((A,t)) = 1/4$

③ $Pr((B,h)) = (1/2)(2/3) = 1/3$

④ $Pr((B,t)) = 1/2 * 1/3 = 1/6$

Define 2 **events**:

① E_1 = "chose coin A"

② E_2 = "outcome is HEAD"

$$\Pr(E_1 \mid E_2) = \frac{\Pr(E_1 \cap E_2)}{\Pr(E_2)} = \frac{\frac{1}{4}}{\frac{1}{4} + \frac{1}{3}} = \frac{3}{7}$$

INDEPENDENT EVENTS

Two events *E* and *F* are independent if and only if:

$$\Pr(E_1 \cap E_2) = \Pr(E)\Pr(F)$$

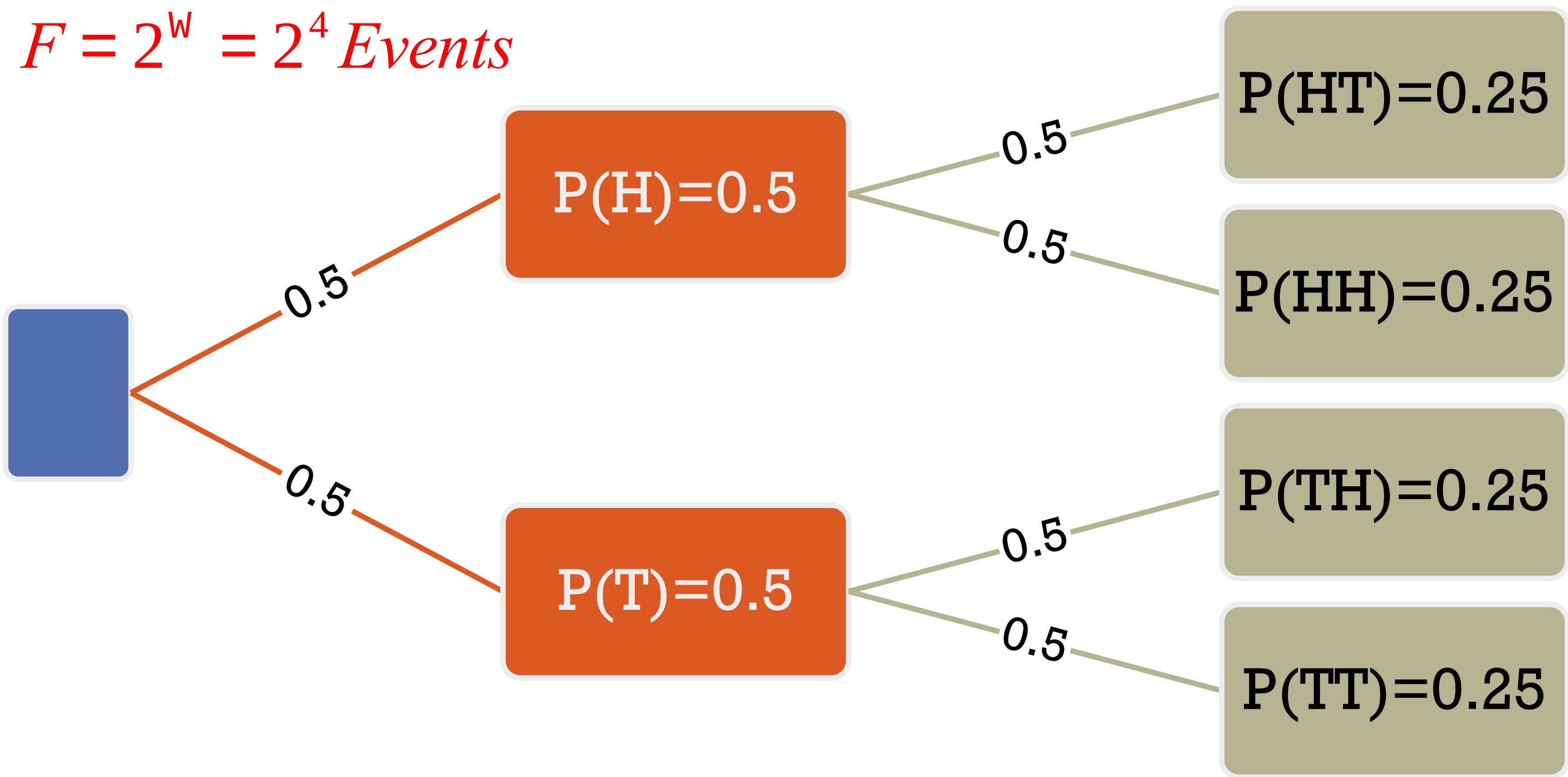
More generally, events $E_1, E_2, \ldots, E_k$ are **mutually independent** if and only if for any subset $I \subseteq [1, k]$

$$\Pr\left(\bigcap_{i \in I} E_i\right) = \prod_{i \in I} \Pr(E_i)$$

EXAMPLE: TOSSING A (FAIR) COIN TWICE

$\Omega = \{HH, HT, TH, TT\}$

$F = 2^{\Omega} = 2^4 \ Events$



P({HH,HT})=P({HT,TT}) = 0.5

IN CLASS EXERCISES

A fair coin was tossed 10 times and always ended up on **HEAD**. What is the likelihood that it will end up **TAIL** next?

IN CLASS EXERCISES

A fair coin was tossed 10 times and always ended up on **HEAD**. What is the likelihood that it will end up **TAIL** next?

The prior observations don't affect the likelihood.
$\rightarrow$ 1/2

IN CLASS EXERCISES

Stan has two kids. One of his kids is a boy. What is the likelihood that the other one is also a boy?

$$E_1 = \text{Two Boys (BB)}$$

$$E_2 = \text{At least one kid is a boy (B)}$$

IN CLASS EXERCISES

Stan has two kids. One of his kids is a boy. What is the likelihood that the other one is also a boy?

E_1 = Two Boys (BB)

E_2 = At least one kid is a boy (B)

$$P(BB) = \frac{1}{2} * \frac{1}{2} = \frac{1}{4}$$

$$P(B) = 1 - P(GG) = 1 - \frac{1}{2} * \frac{1}{2} = \frac{3}{4}$$

$$P(BB|B) = \frac{P(BB \cap B)}{P(B)} = \frac{1/4}{3/4} = \frac{1}{3}$$

BAYESIAN STATISTICS

LAW OF TOTAL PROBABILITY

Theorem: Law of Total Probability

Let $E_1, E_2, \ldots E_n$ be **mutually disjoint** events in a sample space Ω, and $\bigcup_{i=1}^{n} E_i = \Omega$

Then:

$$\boxed{\Pr(B) = \sum_{i=1}^{n} \Pr(B \cap E_i) = \sum_{i=1}^{n} \Pr(B \mid E_i)\Pr(E_i)}$$

BAYES' LAW

Theorem: Bayes' Law

Let $E_1, E_2, \ldots E_n$ be **mutually disjoint** events in a sample space Ω, and $\bigcup_{i=1}^{n} E_i = \Omega$

Then:

$$\boxed{\Pr(E_j \mid B) = \frac{\Pr(B \cap E_i)}{\Pr(B)} \; \frac{\Pr(B \mid E_j)\Pr(E_j)}{\sum_{i=1}^{n} \Pr(B \mid E_i)\Pr(E_i)}}$$

Conditional Probability: $\Pr(A \mid B) = \dfrac{\Pr(A \cap B)}{\Pr(B)}$

Law of Total Probability: $\Pr(B) = \sum_{j=1}^{n} \Pr(B \mid E_j)\Pr(E_j)$

BAYES' LAW

Likelihood

Probability of collecting this data when our hypothesis is true

Prior

The probability of the hypothesis being true before collecting data

$$P(H|D) = \frac{P(D|H)\,P(H)}{P(D)}$$

Posterior

The probability of our hypothesis being true given the data collected

Marginal

What is the probability of collecting this data under all possible hypotheses?

APPLICATION: FINDING A BIASED COIN

- We are given three coins. 2 coins are fair, and the 3^{rd} is biased (landing heads with probability 2/3. We need to identify the the biased coin.
- We flip each of the coins. The first and second come up heads, and the third comes up tails.
- What is the probability that the first coin was the biased one?

APPLICATION: FINDING A BIASED COIN

Let E_i be the event that the *ith* coin flip is the biased one and let B be the event that the three coin flips came up HEADS, HEADS, and TAILS.
Before we flip the coins we have $\Pr(E_i) = 1/3$ for $i=1,\ldots,3$, thus

$$\Pr(B \mid E_1) = \Pr(B \mid E_2) = \frac{2}{3} \times \frac{1}{2} \times \frac{1}{2} = \frac{1}{6}$$

and

$$\Pr(B \mid E_3) = \frac{1}{2} \times \frac{1}{2} \times \frac{1}{3} = \frac{1}{12}$$

Applying Bayes' Law we have...

$$\Pr(E_1 \mid B) = \frac{\Pr(B \mid E_1)(\Pr(E_1)}{\sum_{i=1}^{3} \Pr(B \mid E_i)\Pr(E_i)} = \frac{2}{5}$$

The outcome of the 3 coin flips increases the probability that the first coin is the biased one from 1/3 to 2/5.

IN CLASS EXERCISES

Stan has two kids. One of his kids is a boy. What is the likelihood that the other one is also a boy?

E_1 = Two Boys (BB)

E_2 = At least one kid is a boy (B)

$$P(BB) = \frac{1}{2} * \frac{1}{2} = \frac{1}{4}$$

$$P(B) = 1 - P(GG) = 1 - \frac{1}{2} * \frac{1}{2} = \frac{3}{4}$$

$$P(BB|B) = \frac{P(BB \cap B)}{P(B)} = \frac{1/4}{3/4} = \frac{1}{3}$$

$$P(BB|B) = \frac{P(B|BB) * P(BB)}{P(B)} = \frac{1 * 1/4}{3/4} = \frac{1}{3}$$

IN CLASS EXERCISES: DRUG TEST

- 0.4% of the Rhode Island population use Marijuana*
- Drug Test: The test will produce 99% true positive results for drug users and 99% true negative results for non-drug users.
- If a randomly selected individual is tested positive, what is the probability he or she is a user?

$$P(User|+) = \frac{P(+|User)P(User)}{P(+)}$$

$$= \frac{P(+|User)P(User)}{P(+|User)P(User) + P(+|!User)P(!User)}$$

$$= \frac{0.99 \times 0.004}{0.99 \times 0.004 + 0.01 \times 0.996}$$

$$= 28.4\%$$

SPAM FILTERING WITH NAÏVE BAYES

$$P(spam|words) = \frac{P(spam)P(words|spam)}{P(words)}$$

$$P(spam|viagra, rich, \ldots, friend) = \frac{P(spam)P(viagra, rich, \ldots, friend|spam)}{P(viagra, rich, \ldots, friend)}$$

$$P(spam|words) \approx \frac{P(spam)P(viagra|spam)P(rich|spam)\cdots P(friend|spam)}{P(viagra, rich, \ldots, friend)}$$

WARM UP QUESTION

- Assume a statistical test has a chance of 1% (P=0.01) to be wrong
- How many test can you run before the likelihood of being wrong at least once is 50% or more?

$$P(T_1) = 1 - P(F_1) = 0.99$$

$$P\left(T_1 T_2\right) = 0.99 \times 0.99 \approx 0.98$$

$$P\left(F_1 F_2, F_1 T_2, T_1 F_2\right) = 1 - P\left(T_1 T_2\right) \approx 0.02$$

$$P\left(\text{Being Wrong At Least Once}\right) =$$

$$1 - P\left(T_1 T_2 \cdots T_{n-1} T_n\right) = 1 - P\left(T\right)^n$$

$$n = \log(0.5)/\log(0.99) \approx 68.96 \approx 69$$