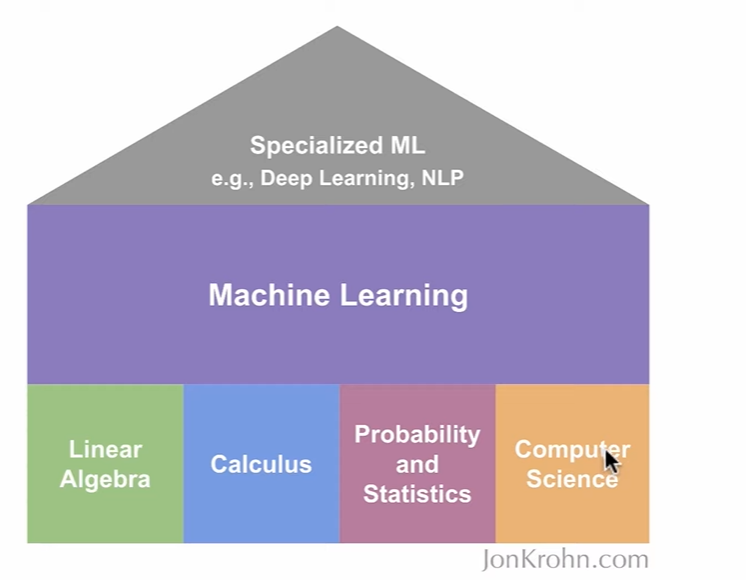
**PROBABILITY & INFORMATION THEORY**

Quantifying uncertainty and building A.I. systems that reason well despite it



1. Intro to Probability

* What probability theory is
* A brief history: Frequentists vs Bayesians
* Application of probability to Machine Learning
* Random variables
* Discrete vs Continuous Variables
* Probability mass and Probability density functions
* Expected Value
* Measures of Central Tendency: Mean, Median, Mode
* Quantiles: Quartiles, Deciles, Percentiles (to get a sense of distribution beyond center)
* The Box-Whisker plot
* Measures of Dispersion: Variance, Standard Deviation, Standard Error
* Measures of Relatedness: Covariance and Correlation
* Marginal and Conditional Probabilities
* Independence and Conditional Independence
* Bayes’ rule

1. Distributions in ML
2. Information theory

A brief history of Probability theory:

* Arab mathematicians (8-13th c.). Largely related to cryptographic communications
* Al-Kindi (9th c.): first known statistical inference
* 16th c.: Italian polymath Gerolamo Cardano
* 17th c.: Frenchmen Pierre de Fermat and Blaise Pascal.
* Modern Probability Theory: Soviet Kolmogorov, Austrian von Mises. Approaches made in this period are frequentist stats, Bayesian stats and Machine Learning

**What Probability Theory is?**

Mathematical study of processes that include *uncertainty*. Probabilities expressed over range of 0 (will not happen) to 1 (will happen). It enables models of future non-deterministic events based on historical data.

*Statistics* allows as to clear quantity in references based on probabilistic events. Provides framework for supporting or rejecting hypotheses.

*Machine learning* is a modeling approach that scales to large, high-dimensional (we have lots of column in our data) data.

*Key concepts:*

*Law of large numbers, random variables, probability distributions, central limit theorem*

*Ω is the set of all possible outcomes*

**Multiple Independent Observations (instead of flipping coin once, let’s do it twice!)**

The main case in here is, one action does not affect the other one. In our example, flipping coin once, does not affect (independent) flipping it the second time.

*So, when flipping a coin there are 2 options we can face: it is either Tails, or it is Heads.*

Let’s imagine we are tossing 2 coins and here probabilities of throwing 2 consecutive heads is:

*(Ω = {HH, HT, TH, TT})*

**Combinatorics**

Combinatorics is a field of mathematics devoted to counting that can be helpful to studying probabilities. We can use factorials (e.g., 4!=4×3×2×1=24 ), which feature prominently in combinatorics, to calculate probabilities instead of painstakingly determining all of the members of the sample space Ω and counting subsets within Ω.

More specifically, we can calculate numbers of outcomes of an event using the “number of combinations” equation:

The left-hand side of the equation is read " 𝑛 choose 𝑘 " and is most quickly understood via an example: If we have three coin flips, 𝑛=3 , and if we're interested in the number of ways to get two head flips (or two tail flips, for that matter), 𝑘=2 . We would read this as "3 choose 2" and calculate it as:

