Probabilistic Graphical Models - Applying the Theory

+======================

# References:

Neapolitan (Learning Bayesian Networks)

Coursera ("Probabilistic Graphical Models", Koller)

Koller/Friedman "Probabilistic Graphical Models"

Also must mention Judea Perl

+====================

# What is a PGM?

PGM is a way to model inherently probabilistic data or system

Graph (Vertices, Edges) to express relationships

Theoretically grounded (Bayes, Conditional Probability)

# Why PGM?

Probabilistic Underpinnings

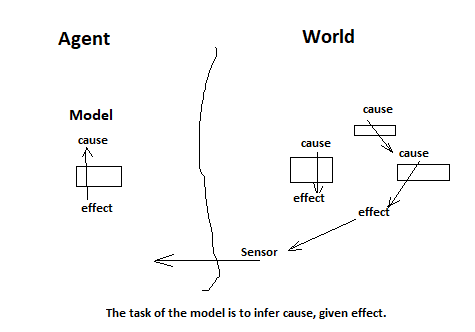
Model-based

## Probabilistic Underpinnings

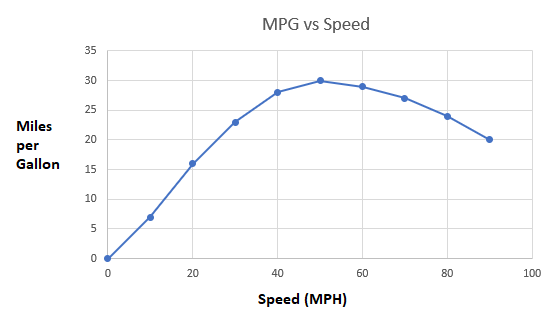
Our agent intends to model something in the real world. We assume the real world is a series of cause-and-effect relationships. We assume each cause-to-effect relationship is a function.

The model receives inputs from the world and that input is (speaking strictly) an effect. It may be the image falling on the retina of the eye, with prior cause-and-effects being lighting, reflection, refraction, occlusion. Or it may be soundwaves received by the ear, the summation of multiple sources, or it may be the tactile sensor receiver resulting from a cold wet surface. But no matter what the source, the sensory input (observed input to our agent) is the effect.

Given effect our agent needs to infer cause. This is the inverse of the original function in the world, and it is a relation. Given an effect there may be more than one cause, and without other context or input there is no way to tell which cause is the one. The only way to predict cause from effect is using probabilistic means. That is why PGMs are so important - the probabilistic underpinnings allow us to accurately model the non-invertible function - cause from effect - the task presented to our agent.



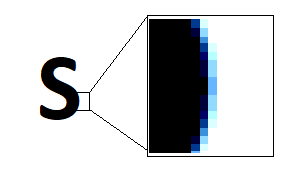
## Example: Miles-per-Gallon (MPG\_ vs Speed



MPG is a function of speed. Given MPG is currently at 25 - what speed is the car traveling? It is likely 35 or 75, but probably not other speeds.

## Example: Portion of a letter

Given the localized portion of an image, what character is this? We see some curve so we know something (it's probably not part of an "I" or "L") but we don't know everything (it might be an "O" or a "P").



## Example: Pinhole camera model

Which basketball is closer? The cause is items 3-d space. The effect is a light received by the retina of the eye (a 2-d image). The cause-to-effect function in the world is 3-d to 2-d. This function is not invertible - you cannot determine distance from a 2-d image. Without other clues we need to model with the information we have. Below we have two basketballs. We can make some assumptions based on size and prior knowledge about basketball standards, and relative position of normal photos in perspective. But our model does not force that - it is possible in the world the smaller ball is in fact closer - it is just predicted as unlikely.





## Example: Counter-top Order Phone

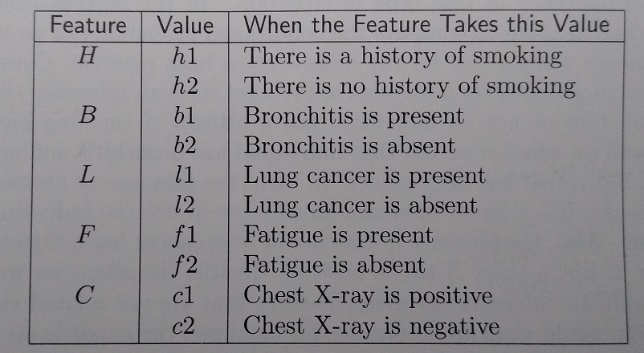
# Model-based (vs Model-free)

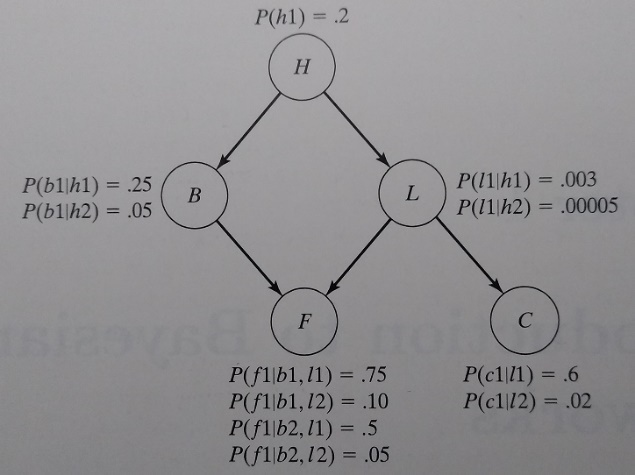
* Our intent is to study how models are established. We are interested in systems that generate models. We are not satisfied with simply generating an accurate model, we want to understand and automate the process. Designing an accurate predictor is not satisfactory.

"Model-free" explicitly advocates that no prescribed model should be used - the system should be 'greenfield' without guidance or external influence. This devalues the study of the processes that are involved.

# Example

Reference: Neapolitan





* Edges represent direct influences
* Lack of edge indicates no direct influence (there may still be correlation)
* Vertices express conditional probability given a combination of parents' values
* A-priori nodes (no parent)
* This is a Bayes (directed) graph. Markov (bidirectional edge) graphs also allowed

# Foundations on the Theory of Probability (Kolmogorov)

1933, Russian. Axiomatic definition of Probability using set theory.

Repeatable Experiment with Outcomes.

Step 1: Define the experiment (set of Outcomes = "Sample Space")

Step 2: Perform the experiment (count Outcomes)

# Terminology:

* Sample Space
* Events
* Elementary Events
* Probability Function
* Random Variable

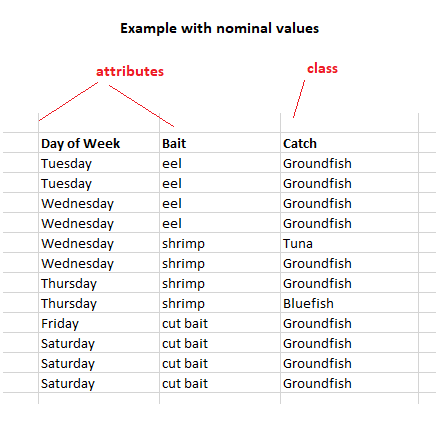
Basics...

* Nominal, Numeric, Ordinal, Parameters, Class

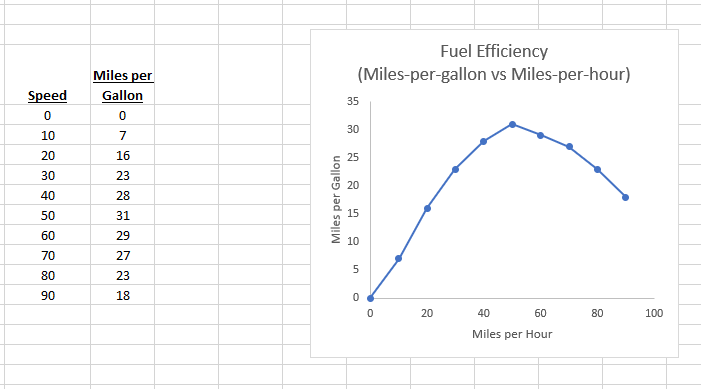
# Basics

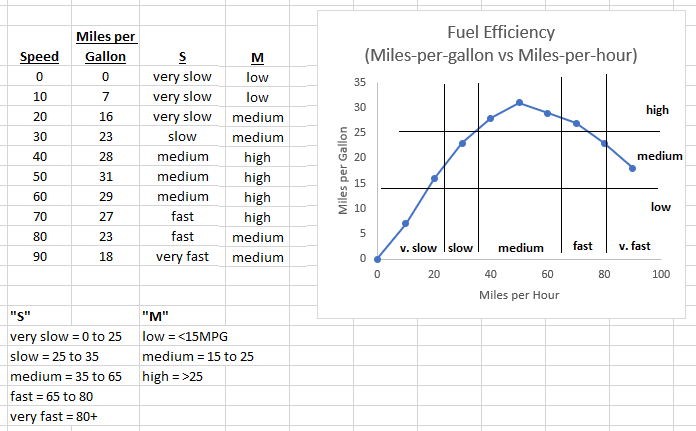
Nominal = categorical

* Numeric = real numbers
* Ordinal = ordered category
  + natural numbers
  + just an ordered list



# Example2 (Numeric Values)





# Tradeoff in Nominal Data Representation

Disadvantages:

* real to nominal conversion introduces loss of accuracy
* Introducing nominal categories increases dimensionality of data

But:

* Allows representing partial (probabilistic) knowledge
* Allows modeling relation (i.e. non-invertible function) such as: **effect given cause**

Review MPG vs Speed: Given

# Defining the Experiment (bottom-up)

Sample space is a set of "elementary events".

An "Event" is a subset of the sample space. An "Elementary Event" is an Event with one element.

# Assigning Probabilities

A "Probability Function" assigns a real number to each Event. Such a probability can be established:

* empirically (sampling)
* subjectively (consulting an "expert")
  + as an expression of historical outcomes
  + as a prediction of future outcomes
  + as a prediction of likelihood of an outcome for a specific future event
* applying Keynes "principle of indifference": (assuming elementary events are equally likely)

Each has advantages/disadvantages.

The pair of Sample Space and Probability Function is called a "Probability Space" and defines the Experiment.

# A Second Way to Define (Top-down)

Two ways to arrive at Sample Space:

1. bottom-up: define the Elementary Events, group into subsets that are Events.
2. top-down: define the Events, enumerate the values. The Sample Space is the cartesian product.

Subjective Probabilities

Fishing

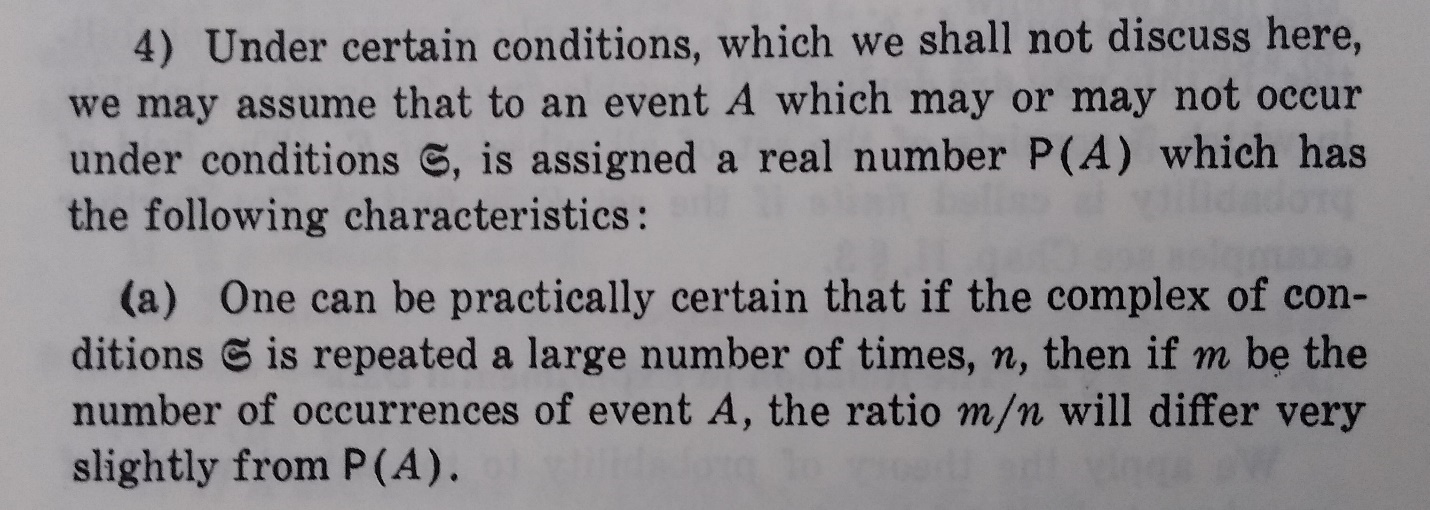
We plan an experiment in ocean fishing. We are interested in the relationship between day of week and type of fish. Our experiment involves pulling a fish from the ocean, classifying its type, and noting if today is weekday or weekend.

We establish two Events:

The sample space is the cross product of Fish and Day. It has 14 elements:

# The Relation to Experimental Data

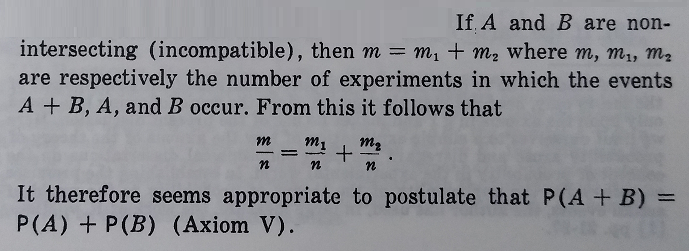
In other words, we establish P(A) such that if the experiment were repeated a large number of times, n, and if m is the number of occurrences of event A, the ratio m/n will approach P(A).



(Kolmogorov Section 2.4.a)

# Conditional Probability and Independence

If we define events A, B to be non-intersecting then total events is the sum of number of individual events, and the probability of either event is the sum of the probabilities of the individual events.



(Kolmogorov Section 2 "The Empirical Deductions of the Axioms")

# Immediate Corollaries: Conditional Probability and Theorem of Bayes

Modern notation:

(Definition of Conditional Probability)

(Bayes Theorem)

# Bringing Intuition to This

Consider the Venn Diagram

# Alternate Means of Definition

(Neapolitan 1.3.1)

Given "some single entity, or set of entities which has features the state of which we wish to determine, but which we cannot determine for certain... In these applications, a random variable represents some feature of the entity being modeled."

For example, if the example considers a set of patients that may or may not have lung cancer, the random variables could be:

ChestXray { positive, negative }

LungCancer { present, absent }

SmokingHistory { yes, no }

After distinguishing the possible values of the random variables, we judge the probabilities of the random variables having their values. Not by prior, nor joint, but rather by relationships, specifically conditional probabilities.

I.e.

P(LungCancer = present

P(ChestXray = positive | LungCancer = absent

P(LungCancer = present | SmokingHistory = yes)

P(LungCancer = present | SmokingHistory = n0)

Obtained from physiaqn, from data, or both.

If the random variable is X, the values of X are the "space" of X.

# Questions One Might Ask

Structure that is imposed by definition of the experiment. I.e. Values of Random Variables.

vs

Structure that comes as a result of sampling. I.e. the sample count is zero.

During the first step - definition of the experiment - is done by establishing random variables and their values. Random Variables are established, each RV has values it can take. Those values are a set, and the the Sample Space of the experiment is the cartesian product of the sets of values of the Random Variables. Values within a Random Variable are disjoint - and by definition cannot co-occur by definition. The experiment inherits this structure.

When Kolmogorov talks about independence, the reference is to counts of events. This is not the fir step (definition) rather it is the second step (execution of the experiment). In other words - events are independent because there is a count of zero. Specifically events A and B are independent if P(A=a0 n B=b0) = 0.

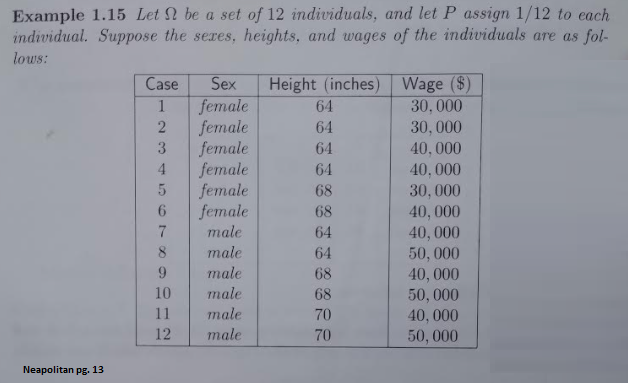
But there is another independence not discussed - that is - p(A=a0, A=a1). Using the Kolmogorov <Define, Perform> sequence this is impossible - Events are disjoint by definition.

In Kolmogorov this definition (structure) is assumed fixed - i.e. the definition is established and not changed.

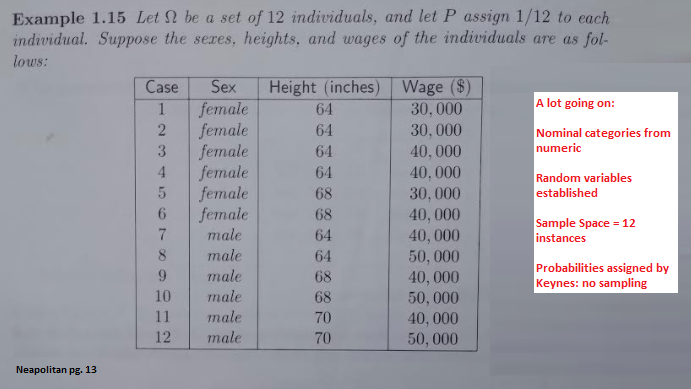
because it falls outside of the possibilities due to the definition of the experiment -

# Neapolitan Page 13 (Random Variables)

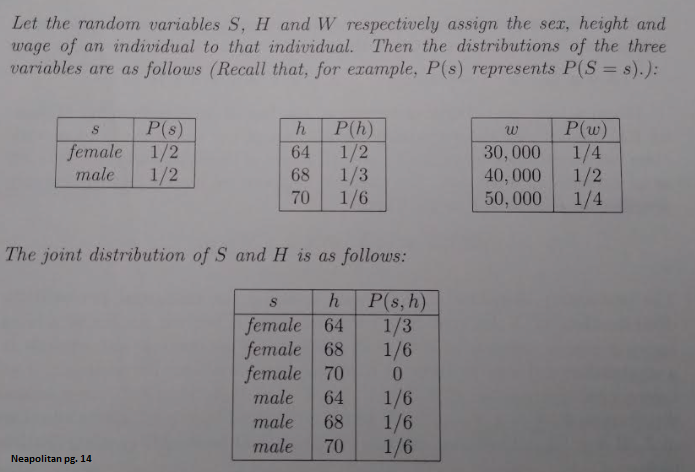
Reference: Neapolitan Page 13



x



x



x

