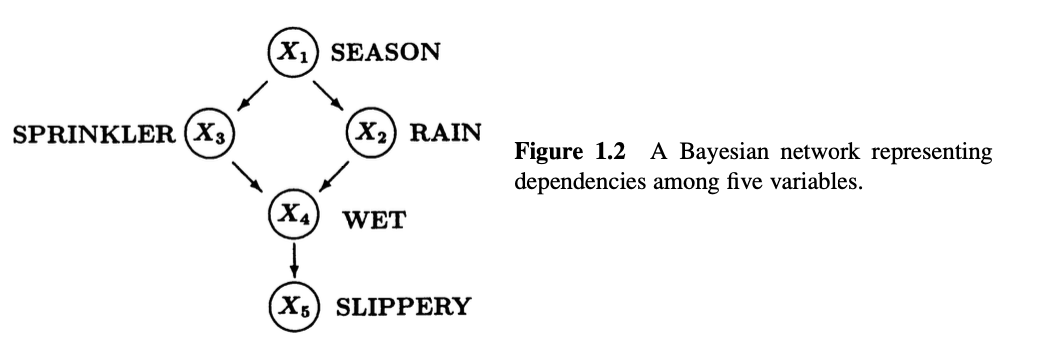
## **Causality (Pearl, J. (2000a). Causality: Models, Reasoning, and Inference. Cambridge University Press, New York. 2nd edition, 2009.** https://mathscinet.ams.org/mathscinet/article?mr=1744773)

**Introduction to Probabilities, Graphs, and Causal Models**

* we say something causes something, but in fact, it is not a direct cause, but it just makes the consequences more likely
  + e.g., "reckless driving causes accidents" or "you will fail the course because you are lazy"
* causal expressions are subject to exceptions
  + "my neighbors roof gets wet when mine gets wet" + "if I hose my roof it will get wet" → my neighbor's roof gets wet whenever I hose mine
* probability theory allows us to focus on causality without having to deal with paradoxes
* we often write probabilities as simply P(A) but when the background information changes, we need to "identify specifically the assumptions that account for our beliefs"
* role of graphs in probabilistic and statistical modeling

1. a convenient way to express assumptions
2. represent joint probability functions
3. facilitate inferences from observations

* in graphical representations, when studying probability of some event X, we only need to look at the Markovian parents of X
  + Markovian parents - minimal set of predecessors that renders X independent of all its other predecessors



* knowing X4 renders X5 independent of {X1, X2, X3}
* d-separation: A path p is said to be d-separated (or blocked) by a set of nodes Z if and only if

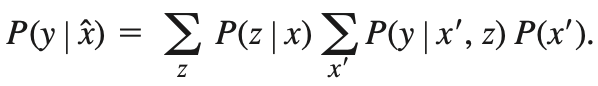
1. p contains a chain i → j or a fork i ← m → such that the middle node m is in Z, oe
2. p contains an inverted fork (or collider) i → m ← j such that the middle node m is not in Z and such that no descendant of m is in Z.

A set Z is said to d-separate X from Y if and only if Z blocks every path from a node in X to a

node in Y.

* in condition 1, two variables become independent of each other one we know the value of the middle variable (vice versa for condition 2)
* X2 and X3 are d-separated from X1, but are not d-separated from {X1, X5}
* first algorithms for probabilistic calculations in Bayesian Networks
  + message passing (limited to trees) - each variable is assigned a simple processor and passes messages asynchronously to its neighbors until convergence
  + join-tree method - network is decomposed into clusters and treat them as a compound variable capable of passing messages
  + cut-set conditioning - "a set of variables is instantiated (given specific values) such that the remaining network forms a tree. The propagation is then performed on that tree, and a new instantiation chosen, until all instantiations have been exhausted; the results are then averaged"
* acyclic graphs allow us to make assumptions about independence, but we cannot infer causality from them – however, they do help with the visualization of the causal effect
* interventions
  + can change a parent-child relationship without changing others
  + to represent an intervention, delete links to the node acted upon, and assign a value to that node
* causal relationships are more stable than probabilistic relationships
  + causal - describe physical relationships in the world
  + probabilistic - reflect what we know or believe about the world
  + as long as no change took place in the environment, causal relationships should not change (pg 25 example)
* traditionally, causal relationships are expression through functional equations
  + a set of equations called structural equation models (pg 27)
  + if each var appears on the left hand side of some distinct equation, then is it called a structural causal model (example pg 27)
* three types of queries: predictions, interventions, and counterfactuals
* Theorem 1.4.1: causal markovian condition
* counterfactuals were previously a challenge, because we don't have the data for it

**Causal Diagrams and the Identification of Causal Effects**

* eelworm example
* front-door criterion: a set of variables X is said to satisfy the front-door criterion relative to an ordered pair of variables (X, Y) if:
  + Z intercepts all directed paths from X to Y
  + there is no unblocked back-door path from X to Z
  + all back-door paths from Z to Y are blocked by X
* if Z satisfies the front-door criterion relative to (X, Y) and P(x, y) > 0, then the causal effect of X on Y is identifiable and 
* do-calculus
  + a causal effect q is identifiable if there exists a finite sequence of transformations that reduces q into a standard probability expression involving observed quantities

**Probability of Causation: Interpretation and Identification**

* probability of necessity - probability that X would not have occurred in the absence of Y, given that X and Y did in fact occur
  + emphasized the absence of alternative processes that are still capable of explaining the effect
  + applications in epidemiology, and legal
* probability of sufficiency - probability that X would have occurred had Y occurred
  + assesses the presence of an active causal process capable of producing the effect
  + applications in policy analysis, AI, psychology
* PN is generally nonidentifiable, due of confounding and sensitivity to the generative processes
* we need both necessary and sufficient causes to explain things
  + basing explanations on generic tendencies or singular-event considerations lead to us omitting crucial information
* formal definitions and equations on pg4 (PN, PS, PNS, PD, PE)
* monotonicity and identifiability under monotonicity and exogeneity(pg9)
* proof pg 12 - if Y is monotonic relative to X, then PNS, PN, and PS are identifiable whenever P(yx) and P(yx') are identifiable
* 9.3 has examples

## **Causal inference in statistics: An overview**

**1 Introduction**

* causal questions are more common that associational, especially in health, social sciences, behavioral sciences
* researchers rarely use the tools to solve causal questions

**2 From association to causation**

**2.1 The basic distinction: coping with change**

* associations between variables can be inferred, but mostly under static conditions
* Causal analysis takes this further, to infer beliefs under both static and dynamic conditions (changing conditions induced by interventions)
* "behind every causal conclusion there must lie some causal assumption that is not testable in observational studies"

**2.2 Formulating the basic distinction**

* a causal concept cannot be defined by a probability distribution alone
* associational: correlation, regression, dependence, conditional independence
* causal: randomization, influence, effect, confounding

**2.3 Ramifications of basic distinction**

* challenges previous statistical literature, such as whether confounding can be given an associational definition (pg 5)
* new notations are needed to express causal relationships
  + need to distinguish between statistical dependence and causal dependence

**2.4 Two mental barriers: untested assumptions and new notation**

* associational assumptions are testable in principle, but causal assumptions cannot be verified without experimental control
* new notation must be unambiguous

**3 Structural models, diagrams, causal effects, and counterfactuals**

* a theory should be able to
  + represent causal questions mathematically
  + have a way to communicate assumptions under which the questions should be answered
  + have a systematic way to answer the question or labeling it unanswerable
  + have a way to determine what new assumptions would make unanswerable questions answerable
* a general theory should not only embrace causal questions, but it should also subsume other theories/methods useful is exploring causation
* Structural causal models combines structural equation models, potential outcome framework, and graphical models for probabilistic reasoning and causal analysis

**3.1 Introduction to structural equation models**

* if X is a disease variable and Y is a symptom of a disease, then Sewall Wright would write y = *βx* + *uy*
  + Nature examines the values of x and u and assigns Y a value according to the equation
  + however, this appears algebraic, but we cannot write *x* = (*y* -*uY*)/*β*, because this can be misinterpreted to mean that the symptoms influence the disease
  + this was later augmented with a "path diagram" (image pg 10)
  + exogenous variables and they represent observed or unobserved background factors (influence but are not influenced by other variables)
  + unobserved variables are often called "disturbances" or "errors", as it is omitted from the model but are determined to be relevant to the other variables in the model
  + dashed double arrow - correlation is presumed possible
  + another of Wright's contribution is being able to write covariances as Cov(X, Y) = β (refer to the diagrams)
* arrows indicate the possibility of a causal relationship, but the lack of a link indicates no influence of one node on the other
* causal assumptions cannot be tested in isolation, but all causal relationships in one model can be tested (using d-separation)
* Barkson's paradox - observations on a common consequence of two independent causes render those causes dependent

**3.2 From linear to nonparametric models and graphs**

* use structural equations, but no need to commit to any specific functional form
* set of functions is *structural* if each function is invariant to possible changes in the form of the other functions (autonomous)
* modeling causal effects and counterfactuals is done through do(x) – the do operator
  + simulates interventions by deleting functions and replacing them with a constant X = x
* P(z, y|do(x0)) is distinct from P(x, y, z) – interventional vs associational
  + if X is treatment, Y is response, and Z is some covariate, then P(z, y|do(x0)) gives proportion of individuals that would attain Y = y, Z = x under situation where X=x0 is administered uniformly to the population
  + can assess efficacy of the treatment by comparing the distribution at different levels of x0
    - average difference: E(Y |do(x′0)) − E(Y |do(x0))
    - risk ratio: E(Y |do(x′0))/E(Y |do(x0))
* causal effect: P(Y= y|do(x)) = ΣzP(z, y|do(x))
  + can P(Y=y|do(x)) be estimated by P(z, x, y) → the problem of *identification*
* identifiability → definition on page 14)
* page 14 - derivation of identifiability
* all causal effects are identifiable when the model is Markovian (the graph is acyclic)
  + non-markovian models are only identifiable under certain conditions that must be determined from the graph structure
* causal markovian condition (pg 15)
  + the truncated factorization of this has us remove all factors associated with the intervened variables → the post-intervention model is Markovian
  + this is not restricted to intervention on a single variable

**3.3 Coping with unmeasured confounders**

* P(Y=y|do(X=x)) = ΣtP(y|t, x)P(t)
  + T is the set of direct causes of X
* we want to select a subset of factors (sufficient set or admissible set) such that if we compare treated vs untreated subjects that have the same values of the selected factors, we get the correct effect of the treatment
* admissible sets (back-door criterion) - a set S is admissible for adjustment if
  + no element of S is a descendant of X
  + the elements of S "block" all "back-door" paths from X to Y (all paths that end with a path pointing to X)
* intuition: back door paths carry spurious associations from X to Y, while paths directed from T to Y carry causative associations. Blocking back door paths ensures that the association is purely causative
* with a sufficient set, we can write P(Y=y|do(X=x), S=s) = P(Y=y|X=x, S\s)

**3.4 Counterfactual analysis in structural models**

* questions of attribution (what fraction of death cases are due to specific exposure) or susceptibility (what fraction of the healthy unexposed population would have gotten the disease had they been exposed) can't be answered in experimental studies
* need probabilistic analysis of counterfactuals (Y would be y had X been x in situation U = u)
* unit level counterfactuals (pg 25)

**3.5 An example: non-compliance in clinical trials**

* four steps

1. define: express the target quantity Q as a function Q(M) that can be computed from any model M
2. Assume: formulate causal assumptions using ordinary scientific language and represent their structural part in graphical form
3. identify: determine if the target quantity is identifiable
4. estimate: estimate the target quantity if it is identifiable, or approximate it, if it is not

**Unit Selection Based on Counterfactual Logic**

* unit selection - aims to identify a set of individuals who are most likely to exhibit a desired mode of behavior (defined in counterfactual terms)

**Introduction**

* two sub-problems
  + evaluation - devise an estimable objective function that would ensure an optimal counterfactual behavior for the selected group
  + search - devise a search algorithm to select individuals based both on their observed characteristics and the above objective function

**Preliminaries**

* *Yx(u) = y* - *Y* would be *y* had *X* been *x* in the unit *U = u*
* *P(Yz=y) = P(Y=y|do(X=x)*
* *P(Yz=y|X=x')* - the probability that Y would be y had X been x if we observed X = x'

**Counterfactual Formulation of the Unit Selection Problem**

* find set of characteristics c that maximizes the benefit associated with the resulting mixture of compilers, defiers, always-takers, and never-takers
* objective function: *argmaxcβP(ra, ra''|c) + 𝜸Pr(ra,ra,|c) + θPr(ra', ra''|c) + δPr(ra',ra'|c)*
* A=a - encouragement; A=a' - otherwise; R=r - positive response; R=r' - otherwise

**Main Results**

* Thm 1: bounds of the benefit function
* Monotonicity - the assumption that a change from X = false to X = true cannot under any circumstances make Y change from true to false
* the benefit of selecting a compiler (β), always-taker(𝜸), never-taker (θ), defier(δ) is said to satisfy gain equality iff β + δ = 𝜸 + θ
* Thm 4: given that Y is monotonic relative to X or gain equality is satisfied, (β - θ)P(yx|z) + (𝜸 - β)P(yx'|z) + θ

## **Explainable AI and Causal Understanding: Counterfactual Approaches Considered**

**Introduction**

* ML models aid us in decision making, but the models are not fully understood – we don't know exactly why they make the decisions that they do
* counterfactual reasoning - how the model inputs need to change in order to yield a different output
* Pearl's work is relevant because it defines causation, which is important because they want to determine if counterfactual approaches identify genuine causal dependence

**Causal Understanding**

* XAI - providing explanations of why machine learning models yield the results that they do
* counterfactual explanations are supposed to provide information about causes, so providing counterfactual explanations are a way to provide information about causation
* basic causal certification - a guarantee that the information provided to the users is always genuine causal information
* complete causal certification - guarantee that the information provided to users is always a complete account of the causal factors that led the model to deliver a particular outcome
* Pearl-Woodward framework can be used to test if an existing approach to XAI passes basic or complete causal certification

**Counterfactual Explanations**

* what are the smallest changes to the model inputs for the output to change?

## **Unit Selection Based on Counterfactual Logic**

* unit selection - aims to identify a set of individuals who are most likely to exhibit a desired mode of behavior (defined in counterfactual terms)

**Introduction**

* two sub-problems
  + evaluation - devise an estimable objective function that would ensure an optimal counterfactual behavior for the selected group
  + search - devise a search algorithm to select individuals based both on their observed characteristics and the above objective function

**Preliminaries**

* *Yx(u) = y* - *Y* would be *y* had *X* been *x* in the unit *U = u*
* *P(Yz=y) = P(Y=y|do(X=x)*
* *P(Yz=y|X=x')* - the probability that Y would be y had X been x if we observed X = x'

**Counterfactual Formulation of the Unit Selection Problem**

* find set of characteristics c that maximizes the benefit associated with the resulting mixture of compilers, defiers, always-takers, and never-takers
* objective function: *argmaxcβP(ra, ra''|c) + 𝜸Pr(ra,ra,|c) + θPr(ra', ra''|c) + δPr(ra',ra'|c)*
* A=a - encouragement; A=a' - otherwise; R=r - positive response; R=r' - otherwise

**Main Results**

* Thm 1: bounds of the benefit function
* Monotonicity - the assumption that a change from X = false to X = true cannot under any circumstances make Y change from true to false
* the benefit of selecting a compiler (β), always-taker(𝜸), never-taker (θ), defier(δ) is said to satisfy gain equality iff β + δ = 𝜸 + θ
* Thm 4: given that Y is monotonic relative to X or gain equality is satisfied, (β - θ)P(yx|z) + (𝜸 - β)P(yx'|z) + θ

## **Causal Intervention for Leveraging Popularity Bias in Recommendation**

**Abstract**

* recommeder systems often face popularity bias issues, which are exacerbated by the data (over-recommending popular items)
* this paper wants to leverage recommender bias to improve recommendation accuracy
  + remove the impact of popularity bias during the training stage
  + inject desired populatity bias during the inference stage

**Introduction**

* use the causal graph to understand how popularity affects the recommendation process
* discovered that popularity is a common cause for interaction probability and whether the item is exposed
* use do-calculus to remove the impact of popularity on whether an item is recommended or not

## **Vaccine side-effects and SARS-CoV-2 infection after vaccination in users of the COVID Symptom Study app in the UK: a prospective observational study**

**Introduction**

* in the UK, the first Covid vaccines were rolled out 12/08/2020 and 01/04/2021
* wanted to investigate side effects and infection rate of vaccination individuals

**Methods**

* used the back-door adjustment to account for differences within the population
* this study was observational, not causal, so they had to remove the causal paths between outcomes and predictors

## **A network theory of mental disorders**

* an important task of psychiatry is to find out where problems come from, and how to solve them
* there is no central disease mechanism for mental disorders like there are for physical illnesses
  + physical illness - experience symptoms, find and remove the cause
* recently, hypothesis that symptoms are not effects of a common cause, but they in fact cause each other
  + delusion may lead to paranoia, which causes social isolation → forms a feedback loop
* can model the interactions between symptoms as a network

**Symptom Networks**

* mental disorders arise from causal interactions between symptoms
* intervention on one symptom changes the probability distribution of other symptoms

**Diagnosis and Treatment**

* which symptoms and present, and which network interactions sustain them

**Mixed Causal Structure Discovery with Application to Prescriptive Pricing**

* prescriptive pricing uses demand modeling and price optimization to maximize future profits
  + demand modeling attempts to reason about price-demand laws by unveiling causal relationships between demands, prices, and objective factors
* wants to describe demand as a causal effect of price, competing prices, promotion, season, etc
* used Pearl's proposal of BNs and SEMs to reason about the causal relationships