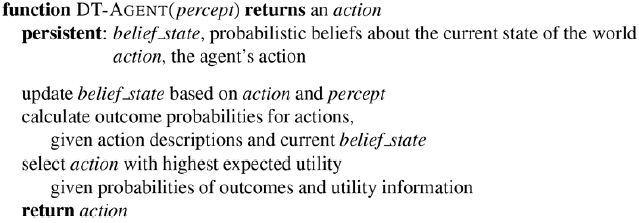
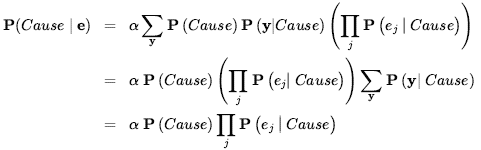
* **How do agents handle uncertainty?**
  + Belief states
  + Contingency plans
  + Drawbacks
    - Most consider every possible belief state, no matter how unlikely
    - Contingency plans try to handle every eventuality no matter how unlikely
    - Sometimes there’s no guaranteed plan to achieve a goal
    - Can’t conclude with absolute certainty
      * Conditions can’t be deduced, so can’t infer a plan succeeds
      * **qualification problem**
        + Needs to qualify every condition for goal
    - Agent can’t guarantee any outcomes, but can provide some degree of belief
* **The rational decision depends on both the relative importance of various goals and the likelihood (and degree to which) they will be achieved**
* Diagnosis is an example of uncertain reasoning
  + But using propositional logic is going to be impossible
    - Rules need to be logically exhaustive
  + Why does logic fail for diagnosis?
    - **Laziness:** Too much work to list the complete set of antecedents or consequents to ensure exceptionless rule (too hard to use such rules)
    - **Theoretical ignorance:** Might not be complete theory or the domain (lots of unknowns)
    - **Practical ignorance:** All necessary tests can’t be run to prove something
* Agent’s knowledge can only provide **degree of belief** so we use **probability theory** instead
  + Instead of a binary outcome, agent considers outcomes as degrees of belief between certainly false and certainly true
* **Probability provides a way of summarizing the uncertainty that comes from our laziness and ignorance**
  + **Probability statements are made with respect to a knowledge state, not with respect to the real world**
* Agent must have preferences among different possible outcomes of various plans
  + **Utility theory** - every state or state sequence has a degree of utility to an agent and the agent will prefer states with higher utility
    - Utility of a state is relative to an agent
    - Utility function can account for any set of preferences
* **Decision theory = probability theory + utility theory**
  + **Principle of maximum expected utility (MEU):** An agent is rational if it chooses the action that yields the highest expected utility averaged over all possible outcomes of the action
  + Decision-theoretic agent’s (DT-AGENT) belief state represents possibilities for world states AND their probabilities



* **Basic probability theory (PT)**
  + **Sample space:** Set of all possible worlds
    - Possible worlds are mutually exclusive and exhaustive
  + Probability models associate numerical probability with each possible world
    - Every possible world has a probability between 0 and 1
    - Total probability of the set of possible worlds is 1
  + **Events**: sets of possible worlds
    - Probability associated with an event is defined to be the sum of the probabilities of the worlds in which it holds
    - PT doesn’t require complete knowledge of probabilities of each world
      * Can constrain underlying model without determining it
  + **Unconditional/Prior probabilities**: degrees of belief in events in the absence of any other information
  + **Conditional**/**Posterior** probabilities: probability of an event based on existing evidence
    - Agent needs to condition on all evidence it’s observed
    - Defined in terms of unconditional probabilities
    - **Product rule**
      * + For *a* and *b* to be true, need *b* to be true, and we also need *a* to be true given *b*
  + **Factored representation**: a possible world is represented by set of variable/value pairs
  + Every random variable is a function that maps from the domain of possible() worlds to some range
    - Variables can be discrete or continuous, finite or infinite
    - **Probability distribution:** assignment of probability for each possible value of random variable
      * **Categorical distribution:** Range is finite and discrete
      * **Probability density function (PDFs)**
        + Can’t use vector for continuous variables (infinite values)
        + Instead define probability as a parameterized function
        + Probabilities are unitless but PDFs are measured with a unit
        + **Joint probability distribution**: probability of a combination of multiple variables
  + **A possible world is defined to be an assignment of values to all the random variables under consideration**
    - Some random variables may be redundant because their values can be obtained in all cases from the values of other variables
    - **Full joint probability distribution**: probability model is determined by joint distribution for all random variables
  + **Inclusion-exclusion principle**
  + If an agent has some degree of a belief in a proposition, the agent should be able to state odds at which it’s indifferent to a bet for or against the proposition
    - If agent 1 expresses a set of degrees of belief that violate axioms of probability theory then there is a combination of bets by agent 2 that guarantees agent 1 will lose money every time
* **Probabilistic inference** - computation of posterior probabilities for query propositions given observed evidence
  + **Marginal probability**: extracting distribution over some subset of variables or single variable
    - Sum up the probabilities for each possible value o the other variables
    - Marginalization Rule
  + Conditioning Rule
    - Apply the product rule to marginalization rule
* Conditional probabilities can be found by obtaining an expression in terms of unconditional probabilities and then evaluating the expression from the full joint distribution
  + - **P(b) is the Normalization constant ():** ensures probabilistic distributions add up to 1
    - Can calculate P(a|b) without normalization constant
      * Just add up probability values to get relative proportions and then normalize each one by dividing by the sum of relative proportions
* General inference procedure
  + Query with single variable *X*
    - E - list of evidence variables
    - e - list of observed values for E
    - Y - remaining unobserved variables
    - Doesn’t scale well in tabular form. For boolean variables it would require space/time O(2n)
    - Full joint distribution is more a theoretical foundation in on which effective approaches can be built
* **(Marginal/Absolute) Independence** 
  + Independence between propositions *a* and *b* can bet written as
    - * All forms are equivalent
  + Independence assertions based usually based on knowledge of domain
    - Reduce amount of info necessary to specify full joint distribution
    - If complete set of variables can be divided into independent subsets, full joint distribution can be factored into separate joint distributions on those subsets
    - Helps reduce size of domain representation and complexity of inference
* **Bayes’ Rule**
  + - Useful because there are many cases where we have good probability estimates for 3 numbers and need to compute the 4th
      * We want to quantify relationship in the diagnostic direction
  + Can avoid assessing prior probability of evidence by computing posterior probability for each value of the query variable and then normalizing the results
  + Bayes’ rule with Normalization
    - * is the normalization constant needed to make entries in sum to 1
  + Diagnostic knowledge is often more fragile than causal knowledge
  + **Conditional independence** given two variables *X* and *Y* given a third variable *Z*
    - **Equivalent forms**
    - Conditional independence can also allow decomposition of the full joint distribution into smaller pieces
      * Size of representation grows as O(n) instead of O(2n)
    - **All conditional independence assertions can allow probabilistic systems to scale up. They are much more commonly available than absolute independence assertions**
* **Naive Bayes Model (Bayesian classifier)**
  + 
    - “Naive” because it’s often used in cases where the effect variables are not strictly independent given the cause variable
    - Works well even when conditional independence assumption is not true
  + 
    - For each possible cause, multiply the prior probability of the cause by the product of the conditional probabilities of the observed effects given the cause, then normalize this result
      * Runtime of calculation is linear in the number of observed effects and doesn’t depend on number of unobserved effects
  + The model is overconfident in its predictions