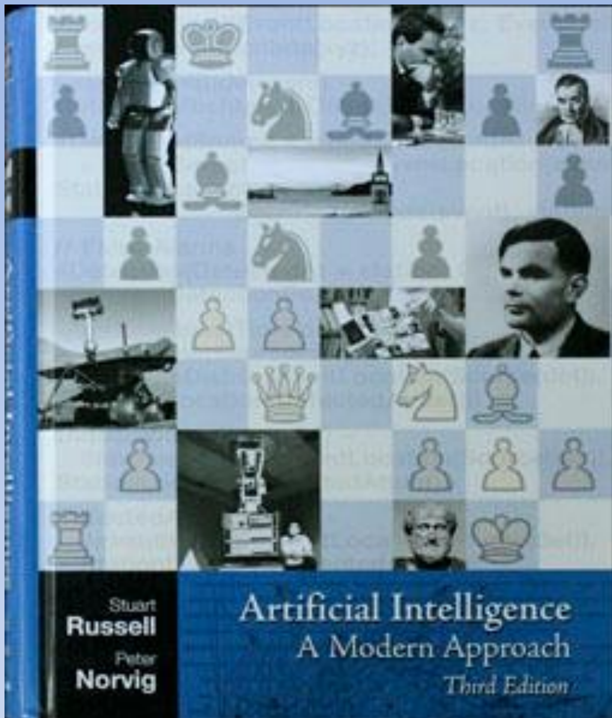
AI Text
Chapter 13



13 QUANTIFYING UNCERTAINTY

In which we see how an agent can tame uncertainty with degrees of belief.

13.1 ACTING UNDER UNCERTAINTY

UNCERTAINTY

Agents may need to handle **uncertainty**, whether due to partial observability, nondeterminism, or a combination of the two. An agent may never know for certain what state it's in or where it will end up after a sequence of actions.

We have seen problem-solving agents (Chapter 4) and logical agents (Chapters 7 and 11) designed to handle uncertainty by keeping track of a **belief state**—a representation of the set of all possible world states that it might be in—and generating a contingency plan that handles every possible eventuality that its sensors may report during execution. Despite its many virtues, however, this approach has significant drawbacks when taken literally as a recipe for creating agent programs:

- When interpreting partial sensor information, a logical agent must consider *every logically possible* explanation for the observations, no matter how unlikely. This leads to impossible large and complex belief-state representations.
- A correct contingent plan that handles every eventuality can grow arbitrarily large and must consider arbitrarily unlikely contingencies.

Why Probabilities?

- Our knowledge of the world is incomplete.
- Complexity of outcomes requires approximation.
- At atomic level the world is Stochastic

- Must deal with:

UNCERTAINTY

Why Probability Theory

- Declarative Representation
  - Propositional Logic
  - First Order Logic
- Representation has clear semantics
- Well developed methods
  - Statistical Mechanics
  - Theoretical Physics/Chemistry
  - Decision Theory (Economics, Psychology,…)
- Great Success in AI/Machine Learning
  - Bayesian Methods
  - Speech Recognition, Text Understanding, Vision
  - Diagnostics

Review Basic Probability Theory/ Everything you need to know (kinda)

- Permutations and Combinations
- Probability Experiments
- Conditional Probabilities
- Distribution Types
  - Binomial, Normal, Exponential
  - Continuous Distributions
- Probability Sampling and Statistics

Where do probabilities come from?

- There is a low probability of light rain in the afternoon.

- Probability here refers to degree of confidence
- Probability Theory deals with formal foundations:
  - Discussing Estimates
  - Rules for estimates

Interpretation of Probabilities

- Frequentist Approach
- Subjectivist Approach
  - Betting Game:
    - One way to attribute belief

Interpretation of Probabilities Frequentist Approach

- Probabilities represent the frequency of events.
  - The probability of an event is the fraction of times it would occur if we repeat the experiment indefinitely.

Interpretation of Probabilities
Frequentist Approach

- Probabilities represent the frequency of events.
- Coin Flip:
  - We repeatedly flip a coin.
  - If we flip the coin 1000 times, how often would it be heads?
  - If we flip the coin 100,000 times?
  - If we flip the coin indefinitely…?
- When this applies well, It’s Clear Semantics!

Interpretation of Probabilities Some Formalism: Space & Events

- Given this Frequentist Perspective, let’s add some formalism.
- Space of Possible Outcomes:
  - Dice roll: $\Omega = \{1, 2, 3, 4, 5, 6\}$
  - Coin flip: $\Omega = \{H, T\}$
- Set of Measurable Events S that we can assign probabilities.
- Each event $\alpha \in S$ is a subset of Ω

2 Coin Example

- $\Omega = \{(H,H), (H,T), (T,H), (T,T)\}$
- Probability of an event where coins match:
  - $\alpha = \{(H,H), (T,T)\}$

- Other examples

Probability Theory
Event Space Requirements

- It contains the empty event ∅,
- It contains trivial event Ω
- It's closed under union
- It's closed under complementation
  - $\alpha \in S$, then so is $\Omega - \alpha$.

Subjectivist Perspective

- Frequentist View doesn't make sense for statements like:
  - The probability of rain tomorrow afternoon is 0.3.
- Subjectivist perspective is probabilities as Subjective Degrees of Belief.
- Betting game:
  - Based on your degree of belief, you should be willing to wager money.
  - If accurate, you win money, else lose money.
  - Belief is rational if it wins money.

Probabilities & Blackjack

- What is the Probability of getting, with the first 2 cards, Blackjack (Natural)

- FIRST:
  - Permutations
  - Combinations

Permutations

- Permutation: An arrangement or selection of objects (without replacement) for which the selection order is important.
  - How many ways can the letters in the word 'CAT' be arranged:
    - ACT, ATC, CAT, CTA, TCA, TAC
    - $3! = 3*2*1$

Permutations

- The number of permutation of n objects taken r at a time is denoted by:
  - $_nP_r = n!/(n-r)! = n(n-1)(n-2)\ldots(n-r+1)$
  - **(n-r)!** is the number of ways to order the remaining items after choosing **r** of them.
- There are nine players on a softball team. How many ways can three of them be chosen to play Left Fielder, Center Fielder, and Right Fielder.
    - $_9P_3 = 9!/6! = 9*8*7 = 504$

Permutations

- A club has 15 members. In how many ways can a president, vice-president, secretary, and treasurer be chosen?

- ${}_{15}P_4$ = 15!/(15-4)! = 15!/11!
  = 15*14*13*12 = 32,760

Permutations versus Combinations

- Consider the following examples:
  1. There are six students in a club. Three will be chosen to go to a convention to represent the club. How many different ways can the three representatives be chosen?
  2. There are six students in a group. Three will be chosen to go to a convention to represent the group, and will be labeled Delegate 1, Delegate 2, Delegate 3. How many different ways can the three delegates be chosen?

Example 2

2. There are six students in a group. Three will be chosen to go to a convention to represent the group, and will be labeled Delegate 1, Delegate 2, Delegate 3. How many different ways can the three delegates be chosen?
   - Here order is important!
   - We could choose Jo, Jose, and Jim to be Delegate 1, 2 and 3, respectively
   - We could also choose Jim, Jo, and Jose to be Delegate 1, 2, and 3, respectively, and this would be a different choice.

Example 1

1. There are six students in a club. Three will be chosen to go to a convention to represent the club. How many different ways can the three representatives be chosen?

- In example 1 order does not matter, unlike example 2.
- Changing the order of the names does not create a new choice.

Combinations

- An arrangement or selection of object (without replacement) in which the order is not important is called a combination.
- Given the softball team with nine players, how many ways can three players be chosen to go to a convention to represent the group.
- ${}_nC_r = {}_nP_r/r! = \frac{n!}{\boldsymbol{r}!(n-r)!} = \binom{n}{r}$
  - ${}_9C_3 = 9!/(3!*6!) = (9*8*7)/(3*2*1)$
  - **r!** is the number of different orderings of the **r** objects.

Back to Probabilities

- Space of Possible Outcomes:
  - Dice roll: $\Omega = \{1, 2, 3, 4, 5, 6\}$
  - Coin flip: $\Omega = \{H, T\}$
- Set of Measurable Events S that we can assign probabilities.
- Each event $\alpha \in S$ is a subset of Ω
- For a set of events S with equally likely outcomes, Probability of an event $\alpha \in S$ is :
  - $P(\alpha) = |\alpha|/|\Omega|$
  - Fraction of total outcomes where event is true.

Blackjack (Natural)

- What is the probability of getting Blackjack (Natural) with first 2 cards?

Blackjack (Natural)

- What is the probability of getting Blackjack (Natural) with first 2 cards?

- There are 52 cards in one deck.
  - There are 4 Aces and 16 face-cards and 10s.
- First calculate all combinations of 52 elements taken 2 at a time: ${}_{52}C_2$ = (52 * 51) / 2 = 1326.
- We combine now each of the 4 Aces with each of the 16 ten-valued cards: 4 * 16 = 64.
- The probability to get a blackjack (natural): 64 / 1326 = **.0483 = 4.83%**.

Poker Hands

- A poker hand is 5 cards dealt from 52.
- How do we calculate the number of possible poker hands?

Poker Hands

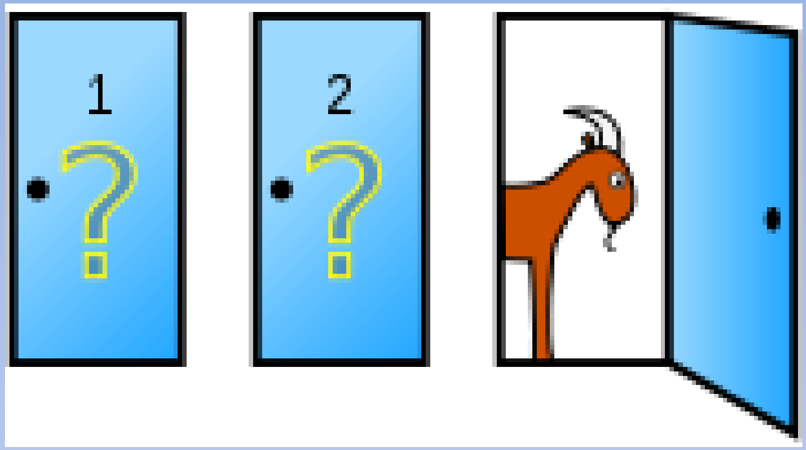
- A poker hand is 5 cards dealt from 52.
- How do we calculate the number of possible poker hands?

- ${}_{52}C_5$
- $\binom{52}{5}$
- 52!/[5!*(52-5)!] = 52*51*50*49*48/5! =
  - 2,598,960

Poker Hands

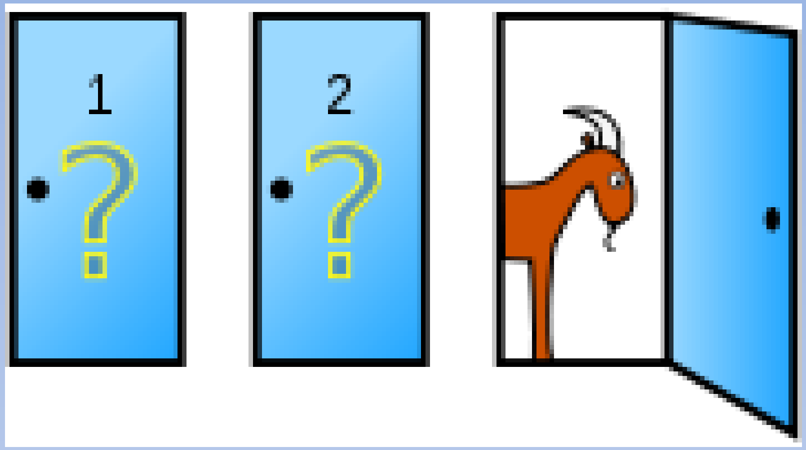
- How do we calculate the Probability of a pair in Poker:
- How many ways to get a pair in Poker
  - $\binom{13}{1}$, choose a face = 13
  - $\binom{4}{2}$, choose two different suits = 6
  - $\binom{12}{3}$, choose remaining card faces = 220
  - $\binom{4}{1}\binom{4}{1}\binom{4}{1}$,Choose remaining suits. 4*4*4 = 64
  - 13*6*220*64 = 1,098,240
- Probability of Pair:
  - 1,098,240/2,598,960 = 42.3%

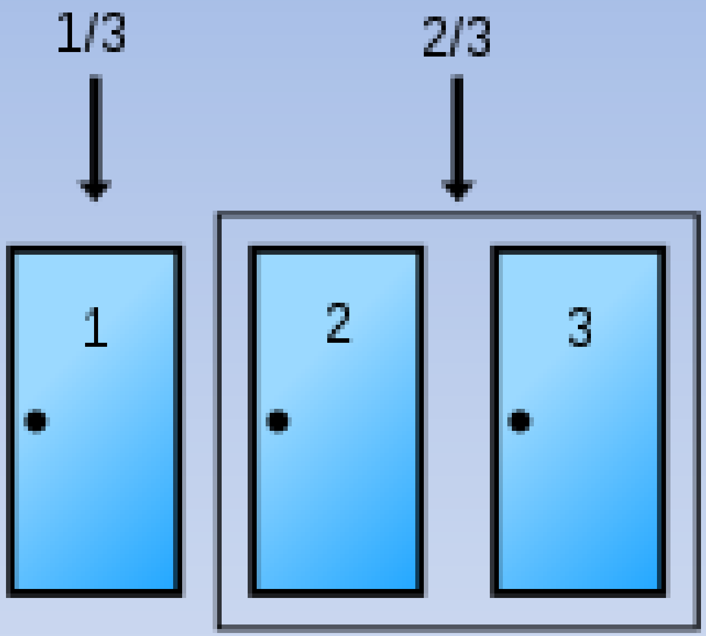
Monty Hall Problem



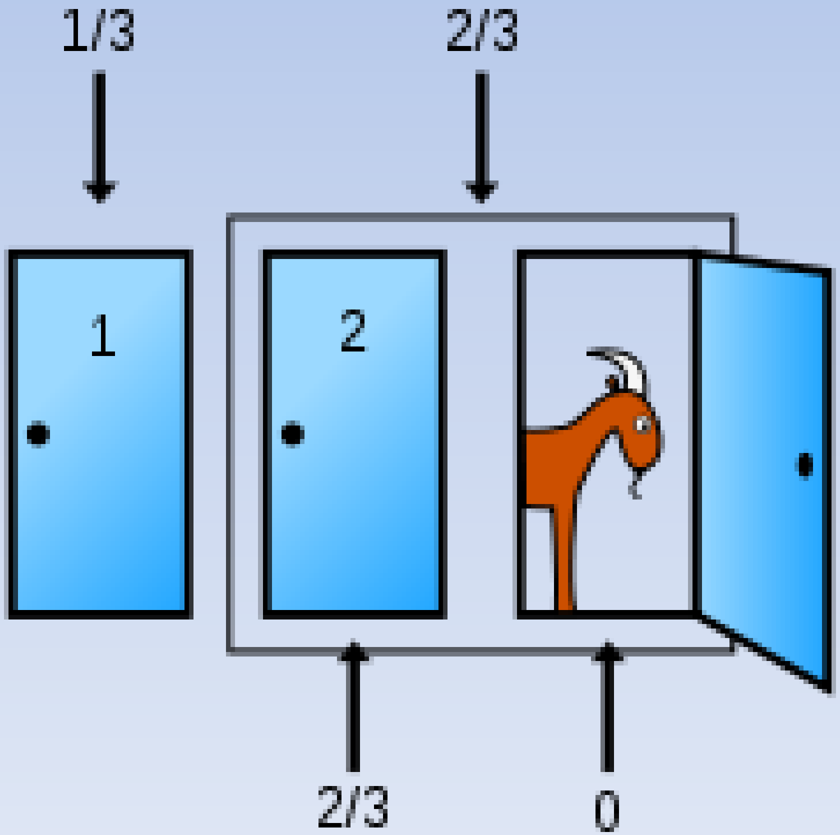
- Suppose you're on a game show, and you're given the choice of three doors:
  - Behind one door is a car;
  - behind the others, goats.
- You pick a door, say No. 1,
- The host, who knows what's behind the doors, opens another door, say No. 3, which has a goat.
- He then says to you, "Do you want to pick door No. 2?“
- Is it to your advantage to switch your choice?

Monty Hall Problem





- Suppose you're on a game show, and you're given the choice of three doors:
  - Behind one door is a car;
  - behind the others, goats.



Probability Rules

- Union Rule: For any events E and F
  - ❖ $P(E \cup F) = P(E) + P(F) - P(E \cap F)$
- If $E \cap F = \emptyset$
  - ❖ E and F are Mutually Exclusive Events
  - ❖ $P(E \cup F) = P(E) + P(F)$
- Complement of any event $E \in S$ is
  - ❖ $E' = S - E$
  - ❖ $P(E') = 1 - P(E)$

Fundamental Counting Principle

- Counting the number of choices when combining groups of items.
- Given two groups one with M possibilities and one with N possibilities.
  - Choose and item from group 1 & 2
  - Total number of choices = MN
- Extends to N groups.

Fundamental Counting Principle
Ordering Pizza

- 3 sizes of pizza (small, medium, large)
- 3 crust choices (thin, thick, regular)
- 6 toppings (beef, sausage, pepperoni, bacon, extra cheese, mushroom)

- How many possible one topping pizzas???

Fundamental Counting Principle
Ordering Pizza

- 3 sizes of pizza (small, medium, large)
- 3 crust choices (thin, thick, regular)
- 6 toppings (beef, sausage, pepperoni, bacon, extra cheese, mushroom)

- How many possible one topping pizzas???
- 3 * 3 * 6 = **54**

Probability Distribution

- Sample Space
  - $S = \{s_1, s_2, \ldots, s_n\}$
- Probabilities
  - $P = \{p_1, p_2, \ldots, p_n\}$
  - p_i is the probability of outcome s_i

Probability Distribution

- A probability distribution P over (Ω, S) is a mapping from events in S to real values that satisfies:
  - $P(\alpha) \geq 0$ for all $\alpha \in S$.
  - $P(\Omega) = 1$.
  - If $\alpha, \beta \in S$ and $\alpha \cap \beta = \emptyset$, then $P(\alpha \cup \beta) = P(\alpha) + P(\beta)$.
- Implied:
  - $P(\emptyset) = 0$
  - $P(\alpha \cup \beta) = P(\alpha) + P(\beta) - P(\alpha \cap \beta)$.

Russell & Norvig Equation 13.1 & 13.2

- 13.1: Given event ω:
  - $0 \leq P(\omega) \leq 1$ for every ω
  - $\sum_{\omega \in \Omega} P(\omega) = 1$
- 13.2: Also For any proposition ϕ, like 'holding(A)'
  - $P(\phi) = \sum_{\omega \in \phi} P(\omega)$
  - The probability of a proposition is the sum of the probabilities for the outcomes where it is true.

Summing to 1

- Very Important: Probability Model for a space of outcomes *must sum* to 1
- If the values do not sum to 1 you do not have probabilities!
- Important later when we define FACTORS
- Important later when we define NORMALIZATION

Conditional Probabilities: Intro

- Suppose a Calculus I class contains 50 students: 35 Juniors (J), 30 CSci Majors (C), and 25 Junior CSci Majors.
  - $n(S) = 50$
  - $n(J) = 35$
  - $n(C) = 30$
  - $n(J \cap C) = 25$
- What is the probability that a student randomly selected from class is a CSci major??

Conditional Probabilities: Intro

- Suppose a Calculus I class contains 50 students: 35 Juniors (J), 30 CSci Majors (C), and 25 Junior CSci Majors.
  - $n(S) = 50$
  - $n(J) = 35$
  - $n(C) = 30$
  - $n(J \cap C) = 25$
- What is the probability that a student randomly selected from class is a CSci major??
  - $n(C)/n(S) = 0.60$

Conditional Probabilities: Question

- A student is randomly selected from the class.
- If we know that the student is a Junior, what is the probability that the student is a Computer Science Major???

Conditional Probability: Definition

- The probability of an event $\beta \in S$ given that we know event $\alpha \in S$ is true is the relative proportion of outcomes satisfying β among these that satisfy α.
- PGM Equation 2.1
  - $P(\beta \mid \alpha) = \dfrac{P(\alpha \cap \beta)}{P(\alpha)}$

Conditional Probabilities: Question

- A student is randomly selected from the class.
- If we know that the student is a Junior, what is the probability that the student is a Computer Science Major???
- $P(C \mid J) = P(C \cap J) / P(J)$
- $(25/50) / (35/50) \approx 0.714$

Conditional Probabilities
Chain Rule

- $P(\beta \mid \alpha) = \dfrac{P(\alpha \cap \beta)}{P(\alpha)}$
- Chain Rule
  - $P(\alpha \cap \beta) = P(\alpha)P(\beta \mid \alpha)$
- General Chain Rule
  - $P(\alpha_1 \cap \ldots \cap \alpha_k) = P(\alpha_1)P(\alpha_2 \mid \alpha_1) \cdots P(\alpha_k \mid \alpha_1 \cap \ldots \cap \alpha_{k-1})$.

Conditional Probabilities
Chain Rule Example

- $P(\alpha \cap \beta \cap \gamma) = P(\alpha)P(\alpha \mid \beta)P(\gamma \mid \alpha \cap \beta)$.
  - $P(\alpha)P(\alpha \mid \beta) = P(\alpha \cap \beta)$
  - $P(\gamma | \alpha \cap \beta) = \dfrac{P(\alpha \cap \beta \cap \gamma)}{P(\alpha \cap \beta)}$

Bayes Rule

- $P(\alpha|\beta) = \dfrac{P(\beta|\alpha)P(\alpha)}{P(\beta)}$
- Note:
  - $\dfrac{P(\beta|\alpha)P(\alpha)}{P(\beta)} = \dfrac{P(\alpha \cap \beta)}{P(\beta)}$

Bayes Rule Significance

- Bayes rule lets use swap the item we Condition On.
  - Disease versus Symptom
- Frequently much easier to model Conditional Probabilities in one direction than the other!

Example

- Suppose that a tuberculosis (TB) skin test is 95 percent accurate.
  - If the patient is TB-infected, then the test will be positive with probability 0.95
  - if the patient is not infected, then the test will be negative with probability 0.95.

- Now suppose that a person gets a positive test result. What is the probability that he is infected?

Example

- Suppose that a tuberculosis (TB) skin test is 95 percent accurate.
  - If the patient is TB-infected, then the test will be positive with probability 0.95
  - if the patient is not infected, then the test will be negative with probability 0.95.

- Now suppose that a person gets a positive test result. What is the probability that he is infected?

- Naive reasoning suggests that if the test result is wrong 5 percent of the time, then the probability that the subject is infected is 0.95.
  - That is, 95 percent of subjects with positive results have TB.

Example w/ Bayes Rule

- Naive reasoning suggests that if the test result is wrong 5 percent of the time, then the probability that the subject is infected is 0.95.
  - That is, 95 percent of subjects with positive results have TB.
- Bayes' rule needs to consider the prior probability of TB infection together with the probability of getting positive test result.

Example w/ Bayes Rule

- Bayes' rule needs to consider the prior probability of TB infection together with the probability of getting positive test result.
- Suppose that 1 in1000 of the subjects who get tested is infected.
  - P(TB) = 0.001
- Probability of a positive test result requires two cases:
  - Case 1: Person has TB
    - P(TB) * P(Positive|TB) = .001 * 0.95 = 0.00095
  - Case 2: Person without TB
    - P(¬TB) * P(Positive|¬TB) = 0.999*0.05 = 0.04995
  - P(Positive) = Case 1 + Case 2 + 0.0509
- P(TB | Positive) = P(TB ∩ Positive)/P(Positive)
  - (0.001*0.95)/0.0509
  - About 2 Percent

Representation of Events

- Events as Sets of Outcomes not so convenient
- Need attributes over Outcomes:
  - Patient tests positive for TB
  - Patient does not have TB

- Introduce: Random Variables

Student Example

- Consider a Population of Students
- Smart: denotes students that are smart.
- GradeA: denotes students that receive an A in class.

- Introduce random variable: Grade
- Grade is defined by a function mapping Outcomes in Ω to values in the set {A, B, C}

Random Variables

- P(GradeA) means P(Grade = A)
- 'Grade = A' means:
  - $\{\omega \in \Omega : f_{Grade}(\omega) = A\}$.

- Intelligence: is another variable in student example
  - P(Smart) means P(Intelligence = high)
  - Intelligence values include {high, low}

Random Variables

- Categorical (discrete):
  - Take on a values from a set of possible values

- Real valued:
  - Take on infinite number of possible values

Multinomial Distribution

- Given a random variable: X
  - Val(X) = {x^1, x^2, …, x^k}
  - |X| = k

- So we have:
  - $\sum_{i=1}^{k} P(X = x_i) = 1$

Bernoulli Distribution

- Given a random variable: X
  - Val(X) = {x^0, x^1} = {false, true)
  - |X| = 2
- So we have:
  - $\sum_{i=1}^{k} P(X = x_i) = 1$

Marginal Distribution

- Now we can define the Probability Distribution over a single random variable like X.

- Marginal Distribution over X: P(X)
  - Defines a probability for each possible $P(X=x^i)$
- For Example w/ Marginal Distribution over Intelligence:
  - P(Intelligence=high)=0.3
  - P(Intelligence=low)=0.7
  - P(Intelligence ∈ {high, low})

Joint Distribution

- Our example included two variables: Intelligence & Grade
- Joint distribution allow events over both variables:
  - P(Intelligence=high, Grade=A)

Joint Distribution

		Intelligence		
		low	high	
Grade	A	0.07	0.18	0.25
	B	0.28	0.09	0.37
	C	0.35	0.03	0.38
		0.7	0.3	

Joint Distribution

		Intelligence		
		low	high	
Grade	A	0.07	0.18	
	B	0.28	0.09	
	C	0.35	0.03	

- Defines probabilities for P(Intelligence, Grade)
- P(Intelligence=high, Grade=A) = ?

Joints & Marginals

		Intelligence		
		low	high	
Grade	A	0.07	0.18	
	B	0.28	0.09	
	C	0.35	0.03	

- P(Intelligence=high) = ?

Joints & Marginals

		Intelligence		
		low	high	
Grade	A	0.07	0.18	0.25
	B	0.28	0.09	
	C	0.35	0.03	
			0.3	1.0

- P(Intelligence=high) = ?

Canonical Outcome Space

		Intelligence		
		low	high	
Grade	A	0.07	0.18	0.25
	B	0.28	0.09	
	C	0.35	0.03	
			0.3	1.0

- $\mathcal{X}$ = {Intelligence, Grade}
- Atomic Outcomes are full assignments to all variables in $\mathcal{X}$
- A set of variables $\mathcal{X} = \{X_1, X_2, \ldots, X_n\}$ and associated values implicitly define a canonical outcome space.

Marginals and Joints

- Generally: $\mathcal{X}$ = {X_1, X_2, …, X_n}

- Marginalize over X_1:
  - $\mathrm{P}(X_1) = \sum_{X_2,\ldots x_n} P(X_1, X_2, \ldots, X_n)$
  - $\mathrm{P}(X_1 = x_1^2) = \sum_{X_2,\ldots,X_n} P(X_1 = x_1^2, X_2, \ldots, X_n)$
- P(Intelligence=high) =

$$\sum_{g \in Grade} P(Intelligence = high, g)$$

Marginals and Joints

		Intelligence		
		low	**high**	
Grade	A	0.07	**0.18**	0.25
	B	0.28	**0.09**	
	C	0.35	**0.03**	
			0.3	1.0

- P(Intelligence=high) = $\sum_{g \in Grade} P(Intelligence = high, g)$ = 0.3

Marginals and Joints

		Intelligence		
		low	high	
Grade	A	0.07	**0.18**	0.25
	B	0.28	**0.09**	
	C	0.35	**0.03**	
			0.3	1.0

- P(Intelligence=high) = $\sum_{g \in Grade} P(Intelligence = high, g)$ = 0.3
- P(Intelligence=low) = $\sum_{g \in Grade} P(Intelligence = low, g)$ = 0.7
- Marginal Distribution from Joint Distribution sums to 1
  – As Required for a Probability Distribution!

Conditionals w/ Random Variables

- Conditional Probabilities extend naturally to random variables
- P(Intelligence|Grade=A)
  - Conditional Distribution over Intelligence given Grade is an A.

Conditionals w/ Random Variables

- Conditional Probabilities extend naturally to random variables
- P(Intelligence|Grade=A)
  - Conditional Distribution over Intelligence given Grade is an A.
- P(X | Y)
  - Conditional Distribution over values for X given values for Y.

Chain Rule & Bayes Rule

- Chain Rule w/ Random Variables:
  - $P(X, Y) = P(X)\, P(Y|X)$
  - $P(X_1, \ldots, X_k) = P(X_1)P(X_2|X_1) \cdots P(X_k|X_1,\ldots,X_{k-1})$

Independence

- In cases where events do not to each other contribute influence.
  - Flipping two coins: Neither coin influences the outcome of the other coin.
  - Rolling a dice 3 times. Values rolled do not influence future values
- Independent Events: Events that do not influence each other.

Independence: Formally

- An event α is independent of event β in P:
  - $P \models (\alpha \perp \beta)$ IF:
    - $P(\alpha \mid \beta) = P(\alpha)$
    - OR $P(\beta) = 0$
- P satisfies $(\alpha \perp \beta)$ if and only if
  - $P(\alpha \cap \beta) = P(\alpha)P(\beta)$
- $(\alpha \perp \beta)$ implies $(\beta \perp \alpha)$

Independence: Examples

- Flipping Coins
  - Easy to believe: $P(C_2|C_1) = P(C_2)$

Conditional Independence

- A little more complex
- A little more common

Conditional Independence: Formally

- An event α is Conditionally Independent of event β given event γ in P:
  - $P \models (\alpha \perp \beta \mid \gamma)$, if $P(\alpha \mid \beta \cap \gamma) = P(\alpha \mid \gamma)$
  - or if $P(\beta \cap \gamma) = 0$.

Conditional Independence w/ Random Variables

- Let X,Y,Z be sets of random variables
- X is conditionally independent of Y given Z in a distribution P if
  - P satisfies $(X = x \perp Y = y \mid Z = z)$ for all values $x \in Val(X), y \in Val(Y)$, and $z \in Val(Z)$.
  - Variables in set Z are often said to be observed.
- Marginally Independent if $(X \perp Y \mid \emptyset)$
  - writen $(X \perp Y)$
  - X and Y are marginally independent.

Conditional Independence

- The distribution P satisfies $(X \perp Y \mid Z)$ if and only if
  - $P(X,Y \mid Z) = P(X \mid Z)P(Y \mid Z)$

Independence & Conditional Independence

- Tractability key idea

- Islands of Tractability!

Probability Queries

- Evidence: subset **E** of random variables in the model, and an instantiation *e* to these.
- Query Variables: a subset **Y** of random variables in the network.

- P(Y | E = e)
  - the posterior probability distribution over the values y of Y, conditioned on the fact E = e.
  - This expression can also be viewed as the marginal over Y, in the distribution we obtain by conditioning on e.

maximum a posteriori *probability* (*MAP*) Query

- High Probability Assignment given Evidence.
- W = $\mathcal{X}$ −E

- MAP(W | e) = argmax_wP(w, e)
  - w is the assignment of values:
    - Highest: (w_1=a_1, w_2=a_2, … w_j=a_j)
    - given values (e_1=b_1, …, e_k=b_k)

MAP Query

Event	Probability
a0	0.4
a1	0.6

Event	Evidence	P(Evidence|Event)
a0	b0	0.1
a0	b1	0.9
a1	b0	0.5
a1	b1	0.5

- MAP(A) = ?

- MAP(A, B) = ?

MAP Query

Event	Probability
a0	0.4
a1	0.6

Event	Evidence	P(Evidence|Event)
a0	b0	0.1
a0	b1	0.9
a1	b0	0.5
a1	b1	0.5

- MAP(A) = a1

- MAP(A, B) = a0, b1

Marginal MAP Query

- MAP queries situation where
  - Variable set = $\mathcal{X} = \{x_1, ..., x_n\}$
  - Evidence set = E = {e1, …, ek} $\in \mathcal{X}$
  - Query set = W = $\mathcal{X}$ – E
- Query Set includes all variables not provided as evidence.
- Query a subset of W?

Marginal MAP

Event	Evidence B	Evidence C	P(A, C | B)
a0	b0	c0	0.01
a0	b0	c1	0.03
a0	b1	c0	0.12
a0	b1	c1	0.24
a1	b0	c0	0.1
a1	b0	c1	0.2
a1	b1	c0	0.1
a1	b1	c1	0.2

- MAP(A | b0) = ?

Marginal MAP

Event	Evidence B	Evidence C	P(A, C | B)
a0	b0	c0	0.01
a0	b0	c1	0.03
a0	b1	c0	0.12
a0	b1	c1	0.24
a1	b0	c0	0.1
a1	b0	c1	0.2
a1	b1	c0	0.1
a1	b1	c1	0.2

- MAP(A|b0) =
  - $\text{argmax}_A[\sum_c P(A,C|b0)]$

Expectation

- Given our probability distribution.
- What outcome is expected?
  - What is the expected value.
- Given a single roll of a fair dice
  - Each value 1 thru 6 has equal likelihood

- Expected value?

Expectation

- Given our probability distribution.
- What outcome is expected?
  - What is the expected value.
- Given a single roll of a fair dice
  - Each value 1 thru 6 has equal likelihood

- Expected value = 3.5

EP[X] = $\sum_{x} xP(x)$

Variance

- $\mathbf{V}\text{ar}_P[X] = \mathbf{E}_P[(X - \mathbf{E}_P[X])^2]$.
  - Expected value over the square of the difference from each variable value and the variable's expected value.

Standard Deviation

- $\mathbf{V}\text{ar}_P[X] = \mathbf{E}_P[(X - \mathbf{E}_P[X])^2]$.
  - Expected value over the square of the difference from each variable value and the variable's expected value.
  - $\mathbf{V}\text{ar}[X] = \mathbf{E}(X^2) - [\mathbf{E}(X)]^2$.
- $\text{Sqrt}(\mathbf{V}\text{ar}_P[X])$ is a normalized measure of "distance" from the expected value of X.

Continuous

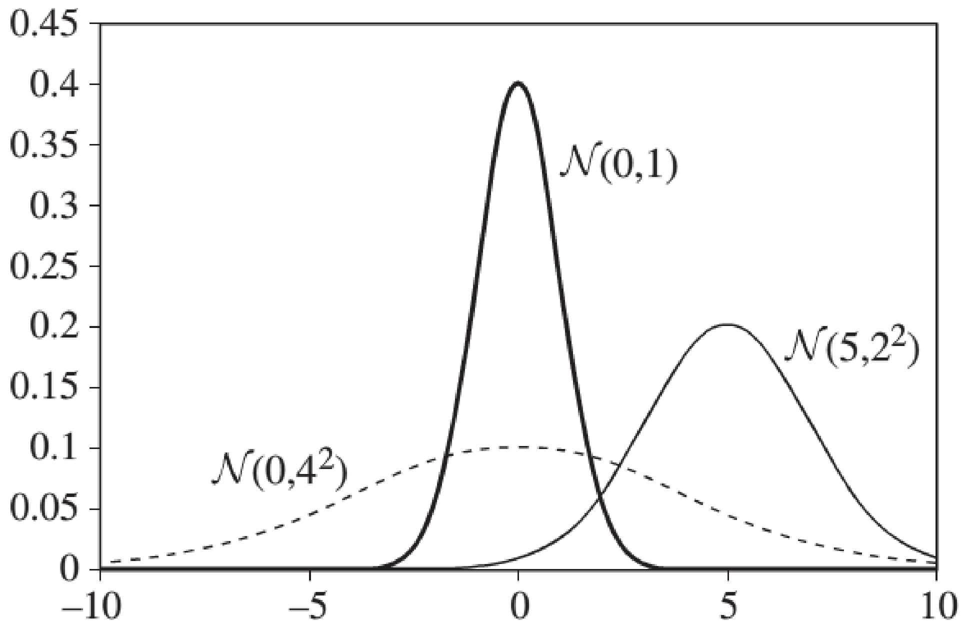


Figure 2.2 Example PDF of three Gaussian distributions

Gaussian/Normal

- $\mathcal{N}(\mu;\sigma^2) =$

$$p(x) = \frac{1}{\sqrt{2\pi\sigma}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}.$$

Standard Deviation

- $\mathbf{V}\mathrm{ar}_P[X] = \mathbf{E}_P[(X - \mathbf{E}_P[X])^2]$.
  - Expected value over the square of the difference from each variable value and the variable's expected value.
  - $\mathbf{V}\mathrm{ar}[X] = \mathbf{E}(X^2) - [\mathbf{E}(X)]^2$.

- Sqrt($\mathbf{V}\mathrm{ar}_P[X]$) is a normalized measure of "distance" from the expected value of X.

Continuous

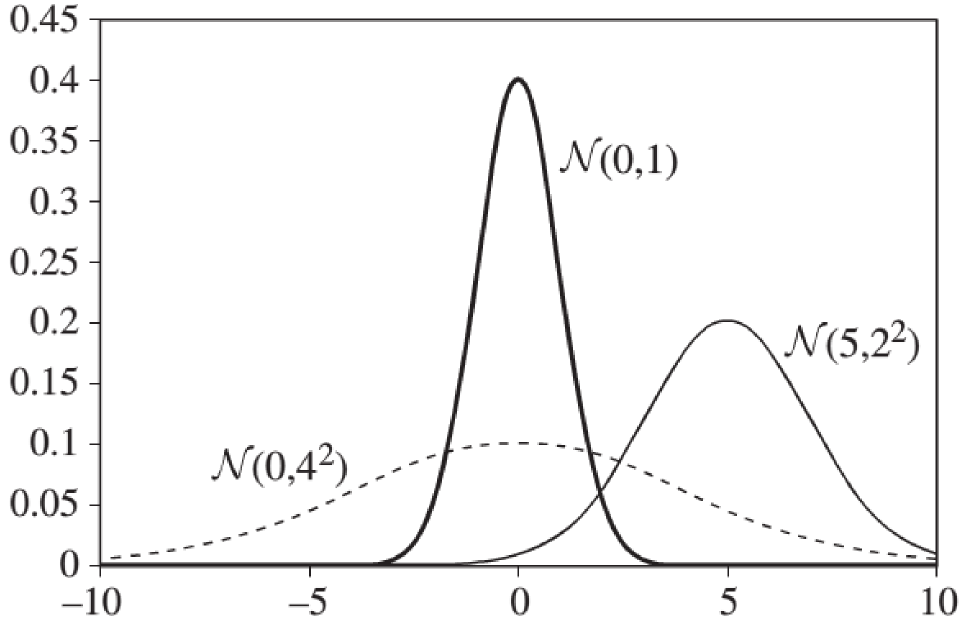


Figure 2.2 **Example PDF of three Gaussian distributions**

Probabilities

- Space of Possible Outcomes:
  - Dice roll: $\Omega = \{1, 2, 3, 4, 5, 6\}$
  - Coin flip: $\Omega = \{H, T\}$
- Set of Measurable Events S that we can assign probabilities.
- Each event $\alpha \in S$ is a subset of Ω
- For a set of events S with equally likely outcomes, Probability of an event $\alpha \in S$ is :
  - $P(\alpha) = |\alpha|/|\Omega|$
  - Fraction of total outcomes where event is true.

Probability Distribution

- Sample Space
  - $S = \{s_1, s_2, \ldots, s_n\}$
- Probabilities
  - $P = \{p_1, p_2, \ldots, p_n\}$
  - p_i is the probability of outcome s_i

Probability Distribution

- A probability distribution P over (Ω, S) is a mapping from events in S to real values that satisfies:
  - $P(\alpha) \geq 0$ for all $\alpha \in S$.
  - $P(\Omega) = 1$.
  - If $\alpha, \beta \in S$ and $\alpha \cap \beta = \emptyset$, then $P(\alpha \cup \beta) = P(\alpha) + P(\beta)$.
- Implied:
  - $P(\emptyset) = 0$
  - $P(\alpha \cup \beta) = P(\alpha) + P(\beta) - P(\alpha \cap \beta)$.

Russell & Norvig Equation 13.1 & 13.2

- 13.1: Given event ω:
  - $0 \leq P(\omega) \leq 1$ for every ω
  - $\sum_{\omega \in \Omega} P(\omega) = 1$

- 13.2: Also For any proposition ϕ, like 'holding(A)'
  - $P(\phi) = \sum_{\omega \in \phi} P(\omega)$
  - The probability of a proposition is the sum of the probabilities for the outcomes where it is true.

Summing to 1

- Very Important: Probability Model for a space of outcomes *must sum* to 1
- If the values do not sum to 1 you do not have probabilities!
- Important later when we define FACTORS
- Important later when we define NORMALIZATION

Student Example

- Intelligence of Student (I)
  - i^0 (low), i^1 (high)
- Difficulty of Course (D)
  - d^0 (easy), d^1 (hard)
- Student's Course Grade (G)
  - g^1 (A), g^2 (B), g^3 (C)

- What's the odds of a smart student getting a B in a difficult Course?
- What's the odds of a smart student getting an A in an easy class??

Student Example
Joint Distrubtion

- Intelligence of Student (I)
  - i^0 (low), i^1 (high)
- Difficulty of Course (D)
  - d^0 (easy), d^1 (hard)
- Student's Course Grade (G)
  - g^1 (A), g^2 (B), g^3 (C)

- How many values needed for our example to cover all possible outcomes (Joint Probability Distribution Table)?

Student Example
Joint Distribution

Intelligence (I)	Difficulty (D)	Grade (G)	P(I, D, G)
I^0	d^0	g^1	0.126
I^0	d^0	g^2	0.168
I^0	d^0	g^3	0.126
I^0	d^1	g^1	0.009
I^0	d^1	g^2	0.045
I^0	d^1	g^3	0.126
I^1	d^0	g^1	0.252
I^1	d^0	g^2	0.0224
I^1	d^0	g^3	0.0056
I^1	d^1	g^1	0.06
I^1	d^1	g^2	0.036
I^1	d^1	g^3	0.024

Intelligence (I)	Difficulty (D)	Grade (G)	P(I, D, G)
I^0	d^0	g^1	0.126
I^0	d^0	g^2	0.168
I^0	d^0	g^3	0.126
I^0	d^1	g^1	0.009
I^0	d^1	g^2	0.045
I^0	d^1	g^3	0.126
I^1	d^0	g^1	0.252
I^1	d^0	g^2	0.0224
I^1	d^0	g^3	0.0056
I^1	d^1	g^1	0.06
I^1	d^1	g^2	0.036
I^1	d^1	g^3	0.024

- What's the probability of a smart student getting a B in a difficult Course?
  - $P(i^1, d^1, g^2) =$

Intelligence (I)	Difficulty (D)	Grade (G)	P(I, D, G)
I^0	d^0	g^1	0.126
I^0	d^0	g^2	0.168
I^0	d^0	g^3	0.126
I^0	d^1	g^1	0.009
I^0	d^1	g^2	0.045
I^0	d^1	g^3	0.126
I^1	d^0	g^1	0.252
I^1	d^0	g^2	0.0224
I^1	d^0	g^3	0.0056
I^1	d^1	g^1	0.06
I^1	**d^1**	**g^2**	**0.036**
I^1	d^1	g^3	0.024

- What's the probability of a smart student getting a B in a difficult Course?
  - $P(i^1, d^1, g^2) = 0.036$
  - Odds = 0.036/(1-0.036) = 0.03734
- Not a satisfying answer… Not quite what we want…

Conditioning: Condition on i^1

Intelligence (I)	Difficulty (D)	Grade (G)	P(I, D, G)
I^0	d^0	g^1	0.126
I^0	d^0	g^2	0.168
I^0	d^0	g^3	0.126
I^0	d^1	g^1	0.009
I^0	d^1	g^2	0.045
I^0	d^1	g^3	0.126
I^1	d^0	g^1	0.252
I^1	d^0	g^2	0.0224
I^1	d^0	g^3	0.0056
I^1	d^1	g^1	0.06
I^1	d^1	g^2	0.036
I^1	d^1	g^3	0.024

Conditioning: Condition on i^1

- Intelligent Students Only

Intelligence (I)	Difficulty (D)	Grade (G)	P(I, D, G)
~~I^0~~	~~d^0~~	~~g^1~~	~~0.126~~
~~I^0~~	~~d^0~~	~~g^2~~	~~0.168~~
~~I^0~~	~~d^0~~	~~g^3~~	~~0.126~~
~~I^0~~	~~d^1~~	~~g^1~~	~~0.009~~
~~I^0~~	~~d^1~~	~~g^2~~	~~0.045~~
~~I^0~~	~~d^1~~	~~g^3~~	~~0.126~~
I^1	d^0	g^1	0.252
I^1	d^0	g^2	0.0224
I^1	d^0	g^3	0.0056
I^1	d^1	g^1	0.06
I^1	d^1	g^2	0.036
I^1	d^1	g^3	0.024

Conditioning: Reduction

- Intelligent Students Only

Intelligence (I)	Difficulty (D)	Grade (G)	P(I, D, G)
I^1	d^0	g^1	0.252
I^1	d^0	g^2	0.0224
I^1	d^0	g^3	0.0056
I^1	d^1	g^1	0.06
I^1	d^1	g^2	0.036
I^1	d^1	g^3	0.024

- But no longer a Probability Distribution!!!
  - Does not SUM TO 1!

CONDITIONING : NORMALIZE

- Intelligent Students Only

Intelligence (I)	Difficulty (D)	Grade (G)	P(I, D, g^1)	Normalized P(D, G | i^1)
I^1	d^0	g^1	0.252	0.63
I^1	d^0	g^2	0.0224	0.056
I^1	d^0	g^3	0.0056	0.014
I^1	d^1	g^1	0.06	0.15
I^1	d^1	g^2	0.036	0.09
I^1	d^1	g^3	0.024	0.06
Total			0.4	1

- But no longer a Probability Distribution!!!
  – Does not SUM TO 1!
- NORMALIZE IT!!!

CONDITIONING : NORMALIZE

- Intelligent Students Only

Intelligence (I)	Difficulty (D)	Grade (G)	$P(I, D, g^1)$	Normalized $P(D, G \mid i^1)$
I^1	d^0	g^1	0.252	0.63
I^1	d^0	g^2	0.0224	0.056
I^1	d^0	g^3	0.0056	0.014
I^1	d^1	g^1	0.06	0.15
I^1	d^1	g^2	0.036	0.09
I^1	d^1	g^3	0.024	0.06
Total			0.4	1

- What's the probability of a smart student getting a B in a difficult Course?
  - $P(g^2 \mid i^1, d^1) = ???$

Let’s condition on i^1, d^1

- Intelligent Students Only

Intelligence (I)	**Difficulty (D)**	**Grade (G)**	**$P(I, D, g^1)$**	**Normalized $P(D, G \mid i^1)$**
I^1	d^0	g^1	0.252	0.63
I^1	d^0	g^2	0.0224	0.056
I^1	d^0	g^3	0.0056	0.014
I^1	d^1	g^1	0.06	0.15
I^1	d^1	g^2	0.036	0.09
I^1	d^1	g^3	0.024	0.06
Total			0.4	1

- What’s the probability of a smart student getting a B in a difficult Course?
  - $P(g^2 \mid i^1, d^1) = ???$

Let’s condition on i^1, d^1

- Intelligent Students in Difficult Classes!

Intelligence (I)	Difficulty (D)	Grade (G)	$P(I, D, g^1)$	Normalized $P(D, G \mid i^1)$
~~I1~~	~~d0~~	~~g1~~	~~0.252~~	~~0.63~~
~~I1~~	~~d0~~	~~g2~~	~~0.0224~~	~~0.056~~
~~I1~~	~~d0~~	~~g3~~	~~0.0056~~	~~0.014~~
I^1	d^1	g^1	0.06	0.15
I^1	d^1	g^2	0.036	0.09
I^1	d^1	g^3	0.024	0.06
Total				

- What’s the probability of a smart student getting a B in a difficult Course?
  - $P(g^2 \mid i^1, d^1) = ???$

Condition on i^1, d^1 : Reduce/Normalize

- Intelligent Students in Difficult Classes!

Intelligence (I)	Difficulty (D)	Grade (G)	$P(I, D, g^1)$	Normalized $P(D, G \mid i^1)$
I^1	d^1	g^1	0.06	0.5
$\mathbf{I^1}$	$\mathbf{d^1}$	$\mathbf{g^2}$	**0.036**	**0.3**
I^1	d^1	g^3	0.024	0.2
Total			0.12	

- What's the probability of a smart student getting a B in a difficult Course?
  - $P(g^2 \mid i^1, d^1) = ???$

Condition on i^1, d^1 : Reduce/Normalize

- Intelligent Students in Difficult Classes!

Intelligence (I)	Difficulty (D)	Grade (G)	$P(I, D, g^1)$	Normalized $P(D, G \mid i^1)$
I^1	d^1	g^1	0.06	0.5
I^1	**d^1**	**g^2**	**0.036**	**0.3**
I^1	d^1	g^3	0.024	0.2
Total			0.12	

- What's the probability of a smart student getting a B in a difficult Course?
  - $P(g^2 \mid i^1, d^1) = 30\%$
  - $P(g^1 \mid i^1, d^1) = 50\%$

Marginalization

- What’s my chance of getting a difficult course???

First :
Full Joint Distribution

Intelligence (I)	Difficulty (D)	Grade (G)	P(I, D, G)
I^0	d^0	g^1	0.126
I^0	d^0	g^2	0.168
I^0	d^0	g^3	0.126
I^0	d^1	g^1	0.009
I^0	d^1	g^2	0.045
I^0	d^1	g^3	0.126
I^1	d^0	g^1	0.252
I^1	d^0	g^2	0.0224
I^1	d^0	g^3	0.0056
I^1	d^1	g^1	0.06
I^1	d^1	g^2	0.036
I^1	d^1	g^3	0.024

Marginalize I, G

Intelligence (I)	Difficulty (D)	Grade (G)	P(I, D, G)
I^0	d^0	g^1	0.126
I^0	d^0	g^2	0.168
I^0	d^0	g^3	0.126
I^0	d^1	g^1	0.009
I^0	d^1	g^2	0.045
I^0	d^1	g^3	0.126
I^1	d^0	g^1	0.252
I^1	d^0	g^2	0.0224
I^1	d^0	g^3	0.0056
I^1	d^1	g^1	0.06
I^1	d^1	g^2	0.036
I^1	d^1	g^3	0.024

Marginalize I, G

Intelligence (I)	Difficulty (D)	Grade (G)	P(I, D, G)
I^0	**d^0**	g^1	0.126
I^0	**d^0**	g^2	0.168
I^0	**d^0**	g^3	0.126
I^0	**D^1**	g^1	0.009
I^0	**D^1**	g^2	0.045
I^0	**D^1**	g^3	0.126
I^1	**d^0**	g^1	0.252
I^1	**d^0**	g^2	0.0224
I^1	**d^0**	g^3	0.0056
I^1	**D^1**	g^1	0.06
I^1	**D^1**	g^2	0.036
I^1	**D^1**	g^3	0.024

Difficulty (D)	P(D)
d^0	0.7
d^1	0.3

Probabilistic Reasoning

- Probability Theory
  - Marginal Probability
  - Conditional Probability

Question

Below is the Joint Probability Distribution Table for the variables: A, B, C, D

A	B	C	D	P(A, B, C, D)
0	0	0	0	0.5
0	0	0	1	0.06
0	0	1	0	0.02
0	0	1	1	0.12
0	1	0	0	0.05
0	1	0	1	0.01
0	1	1	0	0.01
0	1	1	1	0.01
1	0	0	0	0.01
1	0	0	1	0.01
1	0	1	0	0.01
1	0	1	1	0.01
1	1	0	0	0.03
1	1	0	1	0.01
1	1	1	0	0.01
1	1	1	1	0.13

(1) From the table above calculate the following:

(a) **P**(C=0, D=1)

(b) **P**(C=1)

(c) **P**(C=0 | D=1)

(d) Is $C \perp D$?
Why or Why not?

Question
P(C=0, D=1)

A	B	C	D	P(A,B,C,D)
0	0	0	0	0.5
0	0	0	1	0.06
0	0	1	0	0.02
0	0	1	1	0.12
0	1	0	0	0.05
0	1	0	1	0.01
0	1	1	0	0.01
0	1	1	1	0.01
1	0	0	0	0.01
1	0	0	1	0.01
1	0	1	0	0.01
1	0	1	1	0.01
1	1	0	0	0.03
1	1	0	1	0.01
1	1	1	0	0.01
1	1	1	1	0.13

Question:
$\mathbf{P}(C=0, D=1)$

A	B	C	D	P(A,B,C,D)
0	0	0	1	0.06
0	1	0	1	0.01
1	0	0	1	0.01
1	1	0	1	0.01
				=0.09

Question:
P(C=1)

A	B	C	D	P(A,B,C,D)
0	0	0	0	0.5
0	0	0	1	0.06
0	0	1	0	0.02
0	0	1	1	0.12
0	1	0	0	0.05
0	1	0	1	0.01
0	1	1	0	0.01
0	1	1	1	0.01
1	0	0	0	0.01
1	0	0	1	0.01
1	0	1	0	0.01
1	0	1	1	0.01
1	1	0	0	0.03
1	1	0	1	0.01
1	1	1	0	0.01
1	1	1	1	0.13

Question:
P(C=1)

A	B	C	D	P(A,B,C,D)
0	0	1	0	0.02
0	0	1	1	0.12
0	1	1	0	0.01
0	1	1	1	0.01
1	0	1	0	0.01
1	0	1	1	0.01
1	1	1	0	0.01
1	1	1	1	0.13
				=0.32

Question:
P(C=0 | D=1)

A	B	C	D	P(A,B,C,D)
0	0	0	0	0.5
0	0	0	1	0.06
0	0	1	0	0.02
0	0	1	1	0.12
0	1	0	0	0.05
0	1	0	1	0.01
0	1	1	0	0.01
0	1	1	1	0.01
1	0	0	0	0.01
1	0	0	1	0.01
1	0	1	0	0.01
1	0	1	1	0.01
1	1	0	0	0.03
1	1	0	1	0.01
1	1	1	0	0.01
1	1	1	1	0.13

Question:
$\mathbf{P}(C=0 \mid D=1)=P(C=0,D=1)/P(D=1)$

A	B	C	D	P(A,B,C,D)
0	0	0	0	0.5
0	0	0	1	0.06
0	0	1	0	0.02
0	0	1	1	0.12
0	1	0	0	0.05
0	1	0	1	0.01
0	1	1	0	0.01
0	1	1	1	0.01
1	0	0	0	0.01
1	0	0	1	0.01
1	0	1	0	0.01
1	0	1	1	0.01
1	1	0	0	0.03
1	1	0	1	0.01
1	1	1	0	0.01
1	1	1	1	0.13

Question:

P(C=0 | D=1)=P(C=0,D=1)/P(D=1)

A	B	C	D	P(A,B,C,D)
0	0	0	0	0.5
0	0	0	1	0.06
0	0	1	0	0.02
0	0	1	1	0.12
0	1	0	0	0.05
0	1	0	1	0.01
0	1	1	0	0.01
0	1	1	1	0.01
1	0	0	0	0.01
1	0	0	1	0.01
1	0	1	0	0.01
1	0	1	1	0.01
1	1	0	0	0.03
1	1	0	1	0.01
1	1	1	0	0.01
1	1	1	1	0.13

P(C=0 | D=1)
=P(C=0,D=1)/P(D=1)
=0.09/P(D=1)

Question:

P(C=0 | D=1)=P(C=0,D=1)/P(D=1)

A	B	C	D	P(A,B,C,D)
0	0	0	1	0.06
0	0	1	1	0.12
0	1	0	1	0.01
0	1	1	1	0.01
1	0	0	1	0.01
1	0	1	1	0.01
1	1	0	1	0.01
1	1	1	1	0.13
				=0.36

P(C=0 | D=1)
=P(C=0,D=1)/P(D=1)
=0.09/P(D=1)
=0.09/0.36
=1/4=0.25

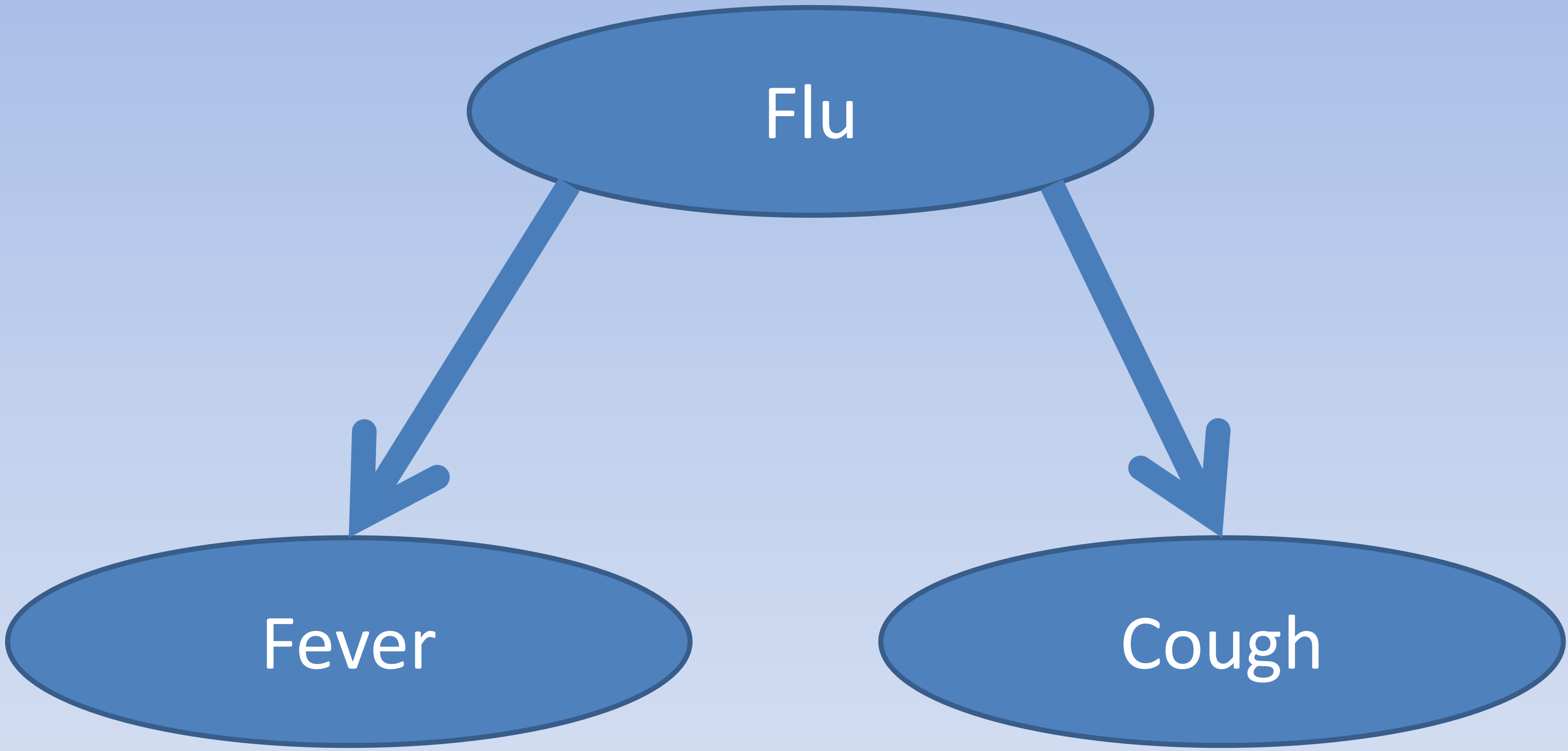
Introducing: Bayesian Networks

- Intuitions similar to Naïve Bayes Model
- Conditional Independencies exploited to allow representation that is Compact & Natural.
- Tailoring allowed so our representation of the distribution only include reasonable independencies!

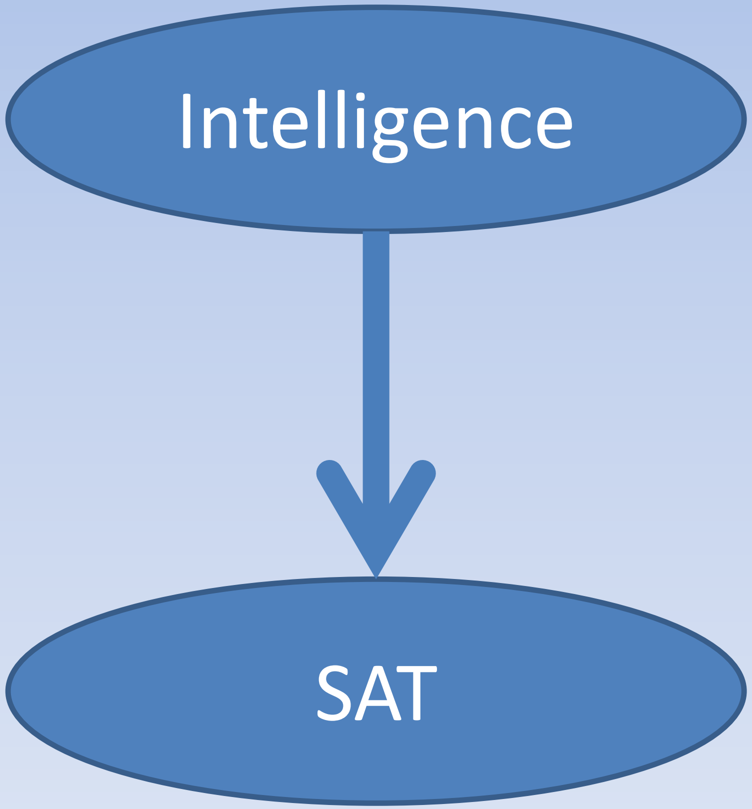
Bayesian Networks : Finally

- Core Idea:
  - Directed Acyclic Graph (DAG)
  - Nodes represent Random Variables in our Domain
  - Edges represent a direct influence from one variable to another.

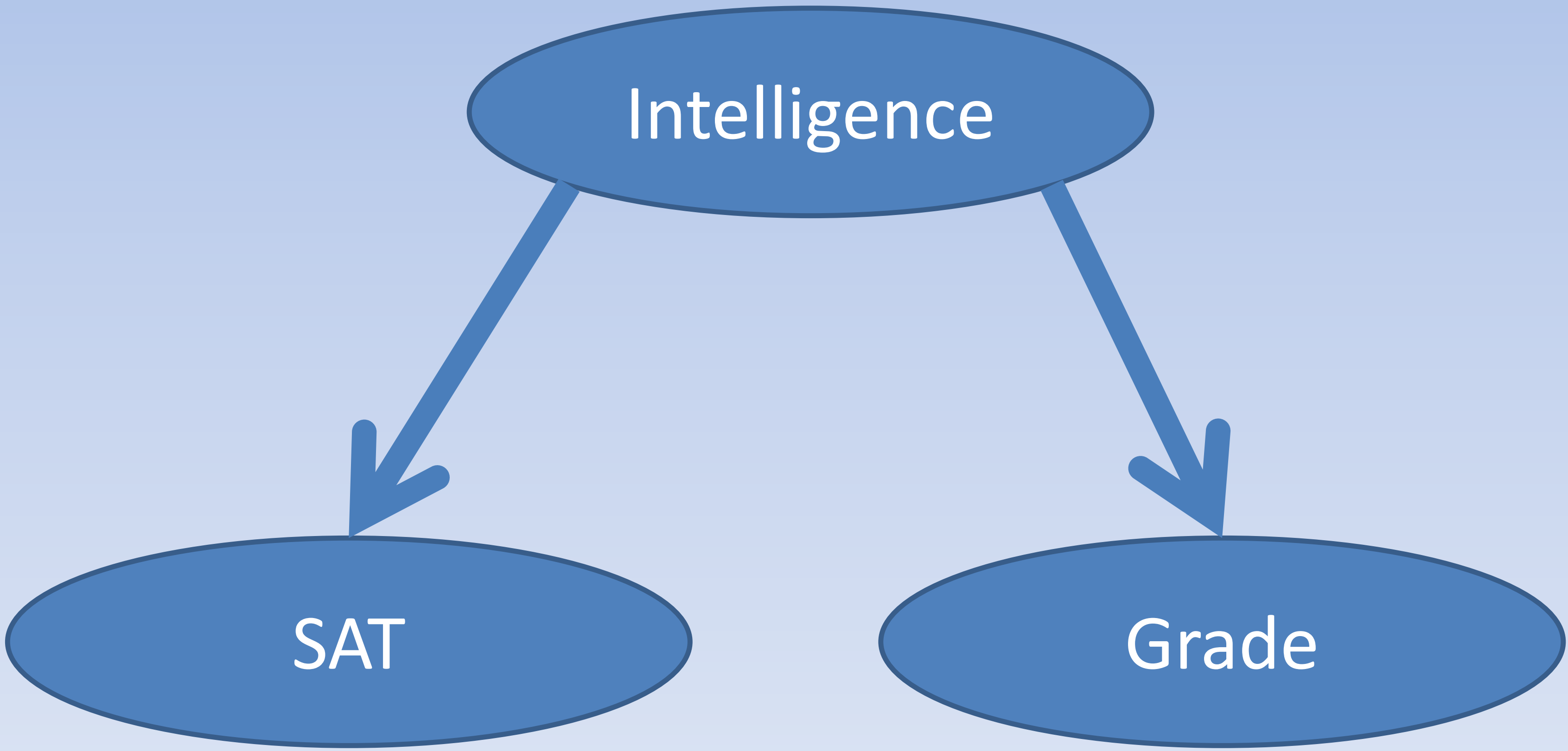
Bayesian Network Graph



Bayesian Network Graph



Bayesian Network Graph



Bayesian Network Graph

- View 1: Data structure that provides the skeleton for representing a joint distribution compactly in a factorized way.

- View 2: Compact representation of Conditional Independence Assumptions.

Bayesian Net Graph

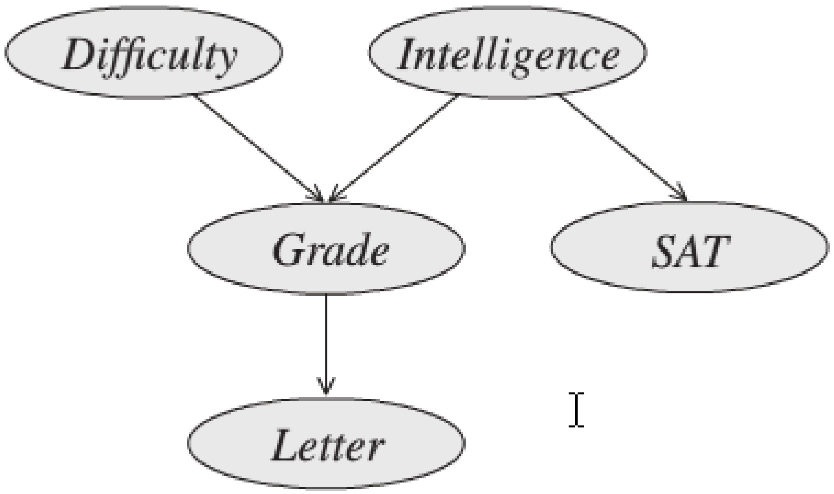


Figure 3.3 The Bayesian Network graph for the Student **example**

Enhanced Example

52 Chapter 3. The Bayesian Network Representation

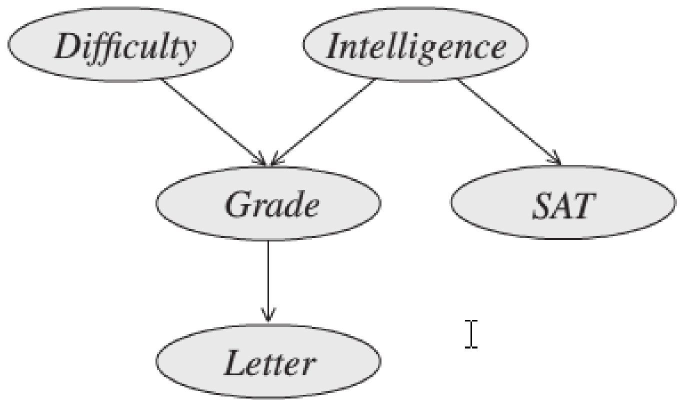
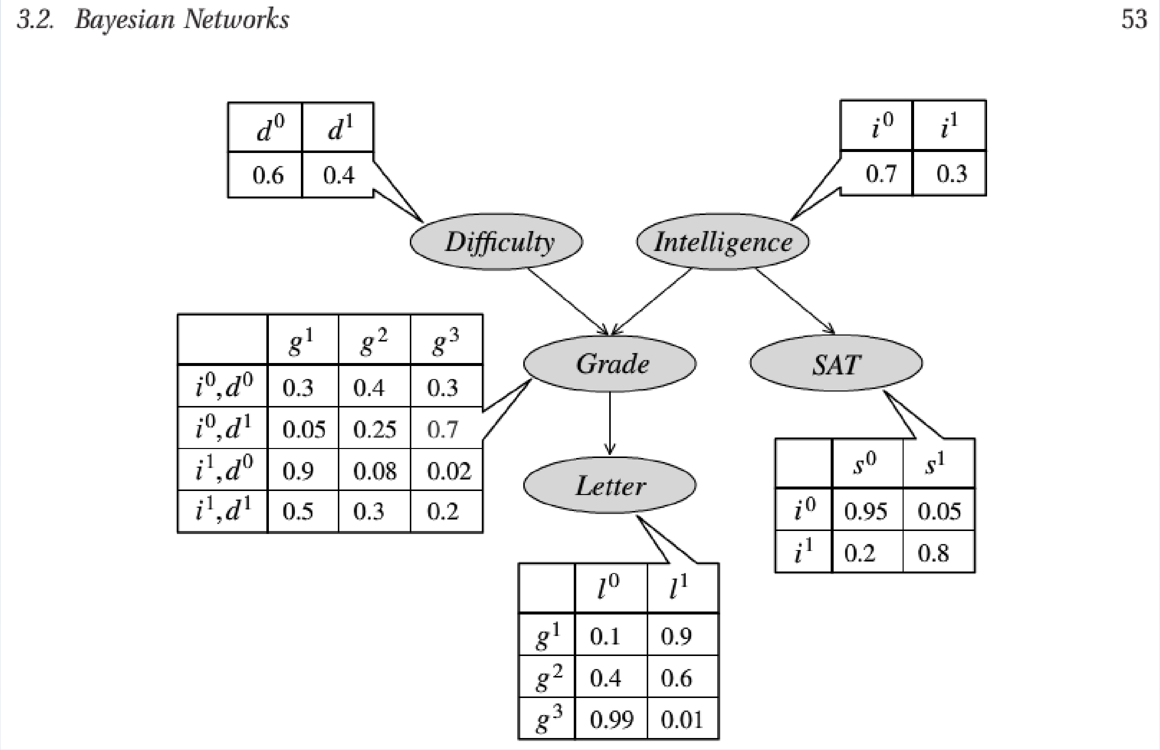


Figure 3.3 **The Bayesian Network graph for the** Student **example**

- Intelligence (I): Val(I)={i^0 (low), i^1 (high)}
- SAT (S): Val(S)={s^0 (low), s^1 (high)}
- Grade (G): Val(G)={g^1 (A), g^2 (B), g^3 (C)}
- ADD:
  - Course Difficulty (D): Val(D)={d^0 (easy), d^1 (hard)}
  - Letter of Recommendation (L): Val(L) = {l^0 (weak), l^1 (strong)}

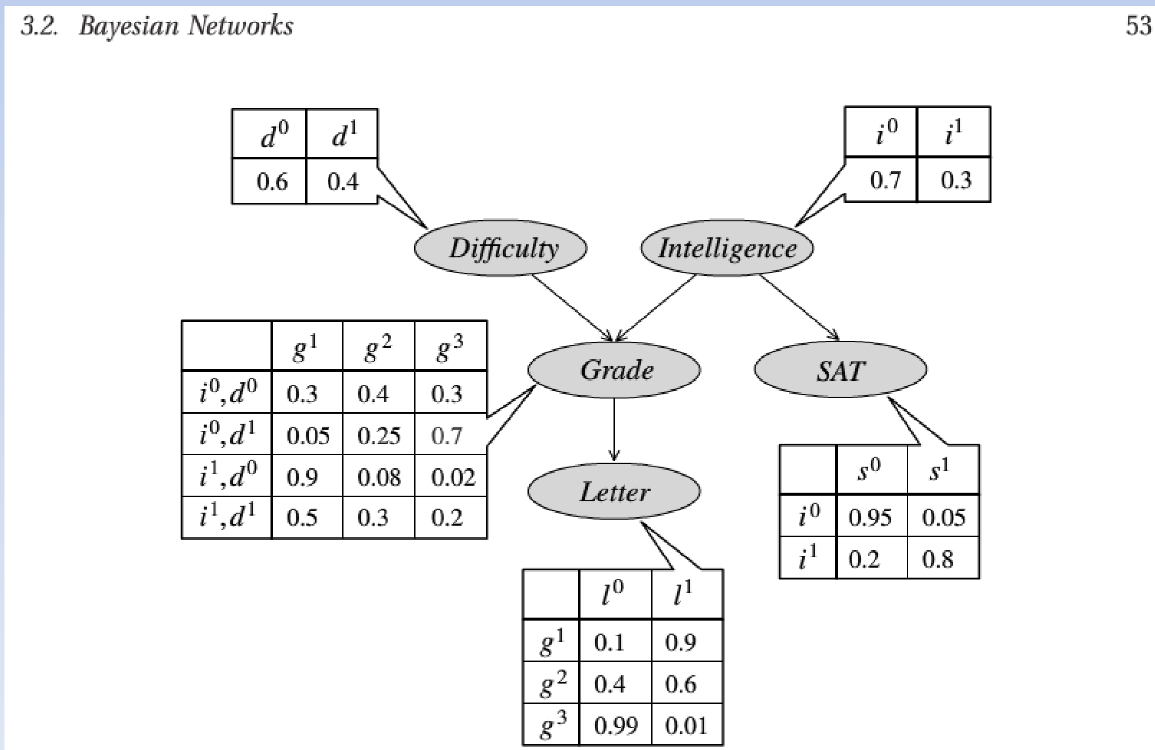
Bayesian Networks : (CPD's)

- 2nd Component of Bayesian Network are Local Probability Models that describe Parent's influence on a Variable.



Bayesian Networks : (CPD’s)

- Each variable is associated with a conditional probability distribution (CPD) that specifies a distribution CPD over the values of X given each possible joint assignment of values to its parents in the model.
- For a node with no parents, the CPD is conditioned on the empty set of variables.



d^0	d^1
0.6	0.4

i^0	i^1
0.7	0.3

	g^1	g^2	g^3
i^0,d^0	0.3	0.4	0.3
i^0,d^1	0.05	0.25	0.7
i^1,d^0	0.9	0.08	0.02
i^1,d^1	0.5	0.3	0.2

	s^0	s^1
i^0	0.95	0.05
i^1	0.2	0.8

	l^0	l^1
g^1	0.1	0.9
g^2	0.4	0.6
g^3	0.99	0.01

Bayesian Network

- The network structure together with its CPDs is a Bayesian network $\mathcal{B}$;
- $\mathcal{B}^{\text{student}}$ refers to the Bayesian network for the student example.
- How do we use $\mathcal{B}^{\text{student}}$ to compute parameters from the full joint distribution?

Let’s Query w/ $\mathcal{B}^{\text{student}}$

- What’s the probability:
  - An intelligent student
  - With High SAT Score
  - Taking an easy class
  - Get’s a B
  - Resulting in a Weak Letter of Recommendation

Let's Query w/ $\mathcal{B}^{\text{student}}$

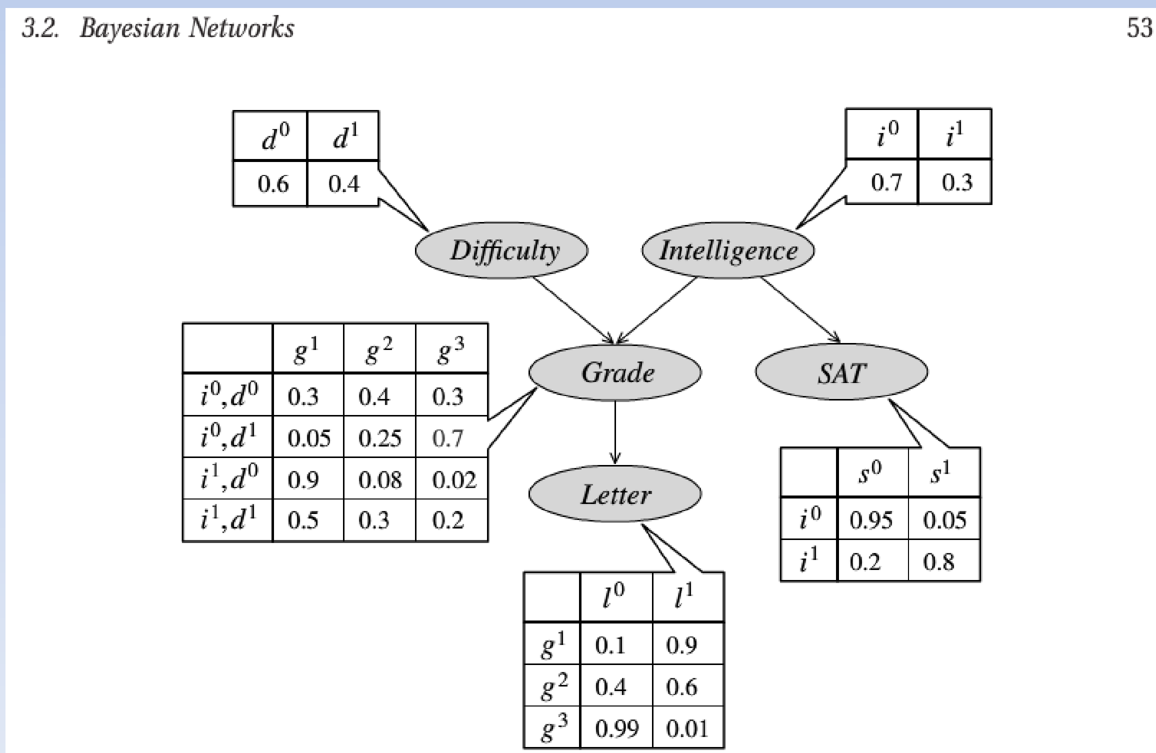
- What's the probability:
  - An intelligent student: $I=i^1$
  - With High SAT Score: $S=s^1$
  - Taking an easy class: $D=d^0$
  - Get's a B: $G=g^2$
  - w/ Weak Letter of Recommendation: $L = l^0$
- $P(i^1, d^0, g^2, s^1, l^0) = ???$

Let's Query w/ $\mathcal{B}^{\text{student}}$

- $P(i^1, d^0, g^2, s^1, l^0) =$
  - $P(i^1)P(d^0)P(g^2 | i^1, d^0)P(s^1 | i^1)P(l^0 | g^2)$

Let's Query w/ $\mathcal{B}^{\text{student}}$

- $P(i^1, d^0, g^2, s^1, l^0) =$
  - ➢ $P(i^1)P(d^0)P(g^2 | i^1, d^0)P(s^1 | i^1)P(l^0 | g^2)$



d^0	d^1
0.6	0.4

i^0	i^1
0.7	0.3

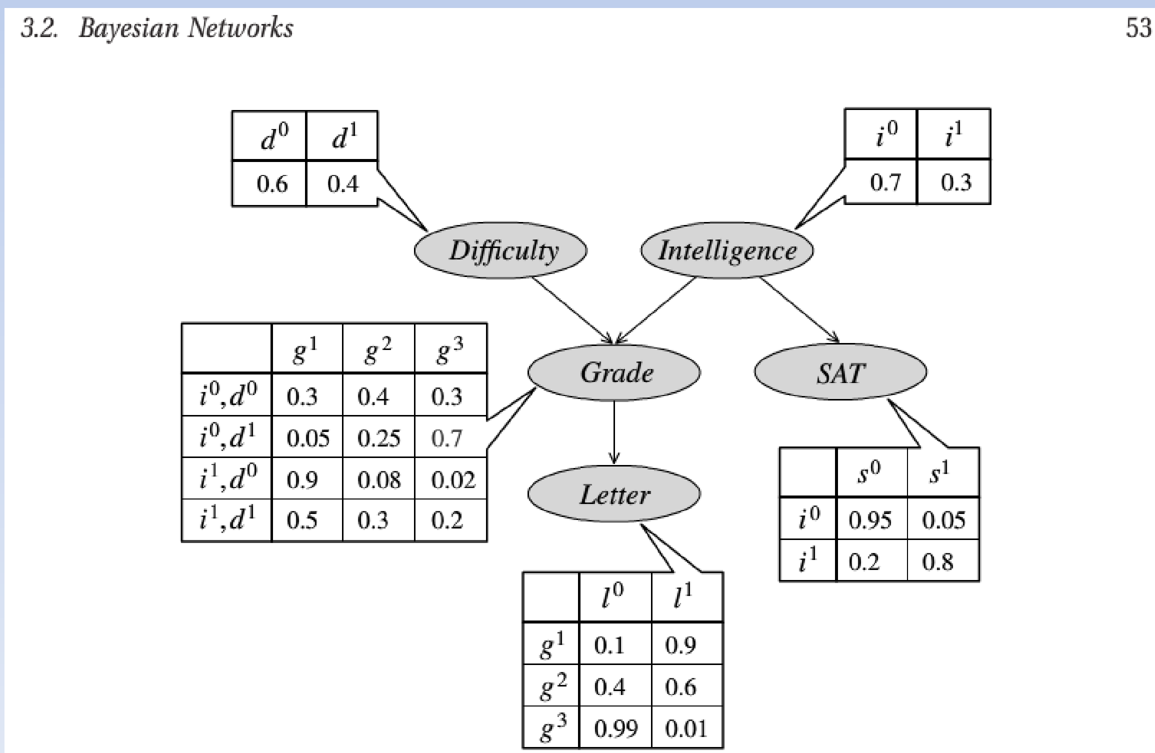
	g^1	g^2	g^3
i^0, d^0	0.3	0.4	0.3
i^0, d^1	0.05	0.25	0.7
i^1, d^0	0.9	0.08	0.02
i^1, d^1	0.5	0.3	0.2

	s^0	s^1
i^0	0.95	0.05
i^1	0.2	0.8

	l^0	l^1
g^1	0.1	0.9
g^2	0.4	0.6
g^3	0.99	0.01

Let’s Query w/ $\mathcal{B}^{\text{student}}$

- $P(i^1, d^0, g^2, s^1, l^0) =$
  - $P(i^1)P(d^0)P(g^2 | i^1, d^0)P(s^1 | i^1)P(l^0 | g^2)$
  - 0.3



d^0	d^1
0.6	0.4

i^0	i^1
0.7	0.3

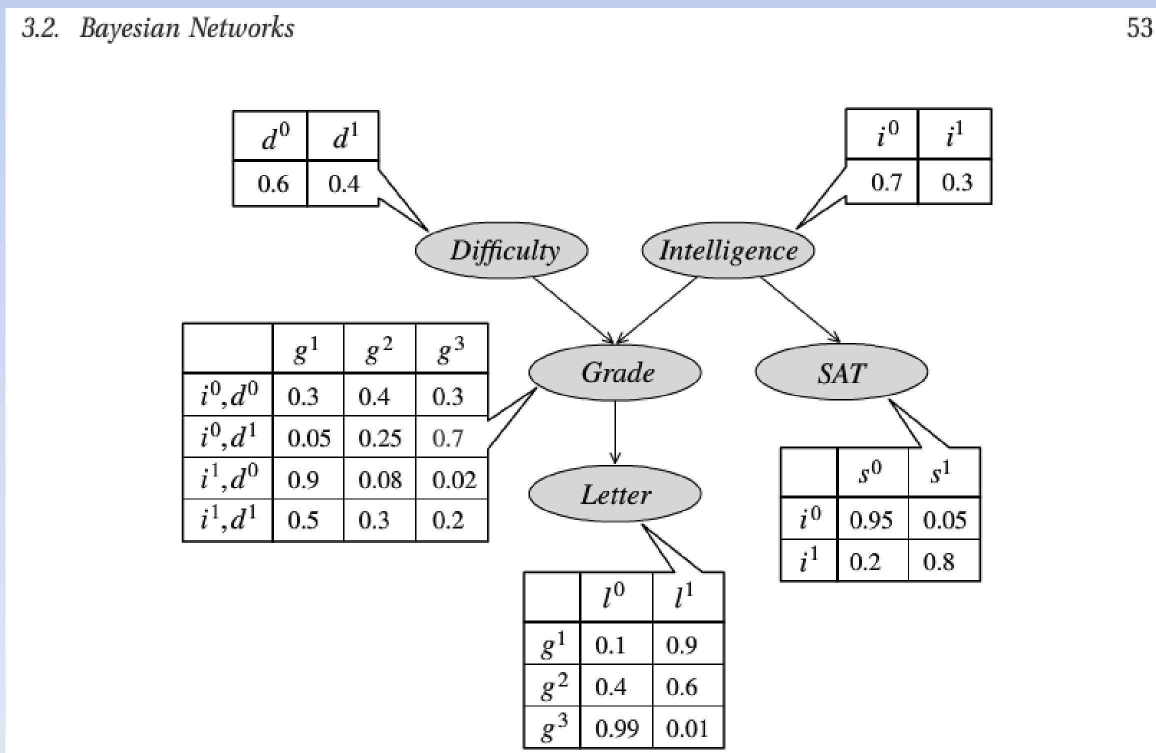
	g^1	g^2	g^3
i^0, d^0	0.3	0.4	0.3
i^0, d^1	0.05	0.25	0.7
i^1, d^0	0.9	0.08	0.02
i^1, d^1	0.5	0.3	0.2

	s^0	s^1
i^0	0.95	0.05
i^1	0.2	0.8

	l^0	l^1
g^1	0.1	0.9
g^2	0.4	0.6
g^3	0.99	0.01

Let’s Query w/ $\mathcal{B}^{\text{student}}$

- $P(i^1, d^0, g^2, s^1, l^0)$ =
  - ➢ $P(i^1)P(d^0)P(g^2 | i^1, d^0)P(s^1 | i^1)P(l^0 | g^2)$
  - ➢ $0.3 \cdot 0.6$


3.2. *Bayesian Networks* 53

d^0	d^1
0.6	0.4

i^0	i^1
0.7	0.3

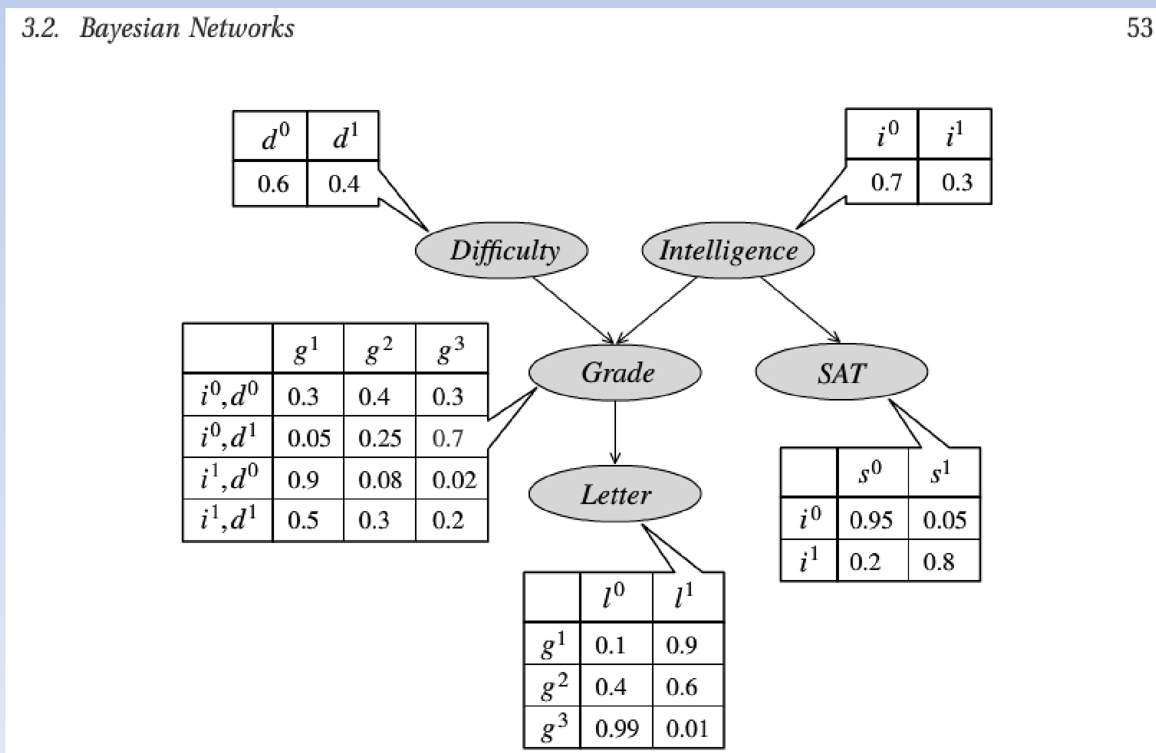
	g^1	g^2	g^3
i^0,d^0	0.3	0.4	0.3
i^0,d^1	0.05	0.25	0.7
i^1,d^0	0.9	0.08	0.02
i^1,d^1	0.5	0.3	0.2

	s^0	s^1
i^0	0.95	0.05
i^1	0.2	0.8

	l^0	l^1
g^1	0.1	0.9
g^2	0.4	0.6
g^3	0.99	0.01

Let's Query w/ $\mathcal{B}^{\text{student}}$

- $P(i^1, d^0, g^2, s^1, l^0) =$
  - $P(i^1)P(d^0)P(g^2 | i^1, d^0)P(s^1 | i^1)P(l^0 | g^2)$
  - $0.3 \cdot 0.6 \cdot 0.08$



d^0	d^1
0.6	0.4

i^0	i^1
0.7	0.3

	g^1	g^2	g^3
i^0, d^0	0.3	0.4	0.3
i^0, d^1	0.05	0.25	0.7
i^1, d^0	0.9	0.08	0.02
i^1, d^1	0.5	0.3	0.2

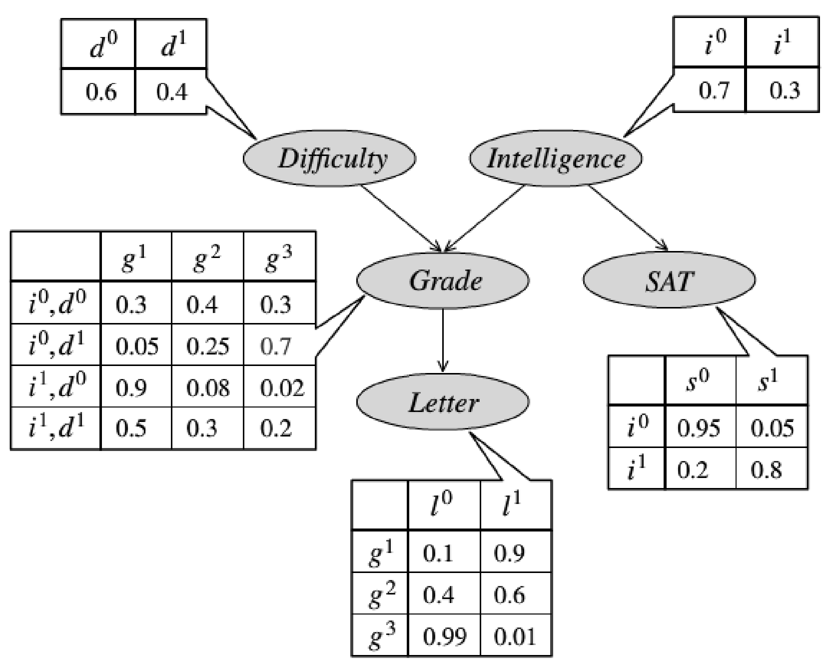
	s^0	s^1
i^0	0.95	0.05
i^1	0.2	0.8

	l^0	l^1
g^1	0.1	0.9
g^2	0.4	0.6
g^3	0.99	0.01

Let's Query w/ $\mathcal{B}^{student}$

- $P(i^1, d^0, g^2, s^1, l^0) =$
  - $P(i^1)P(d^0)P(g^2 | i^1, d^0)P(s^1 | i^1)P(l^0 | g^2)$
  - $0.3 \cdot 0.6 \cdot 0.08 \cdot 0.8$

3.2. *Bayesian Networks* 53



d^0	d^1
0.6	0.4

i^0	i^1
0.7	0.3

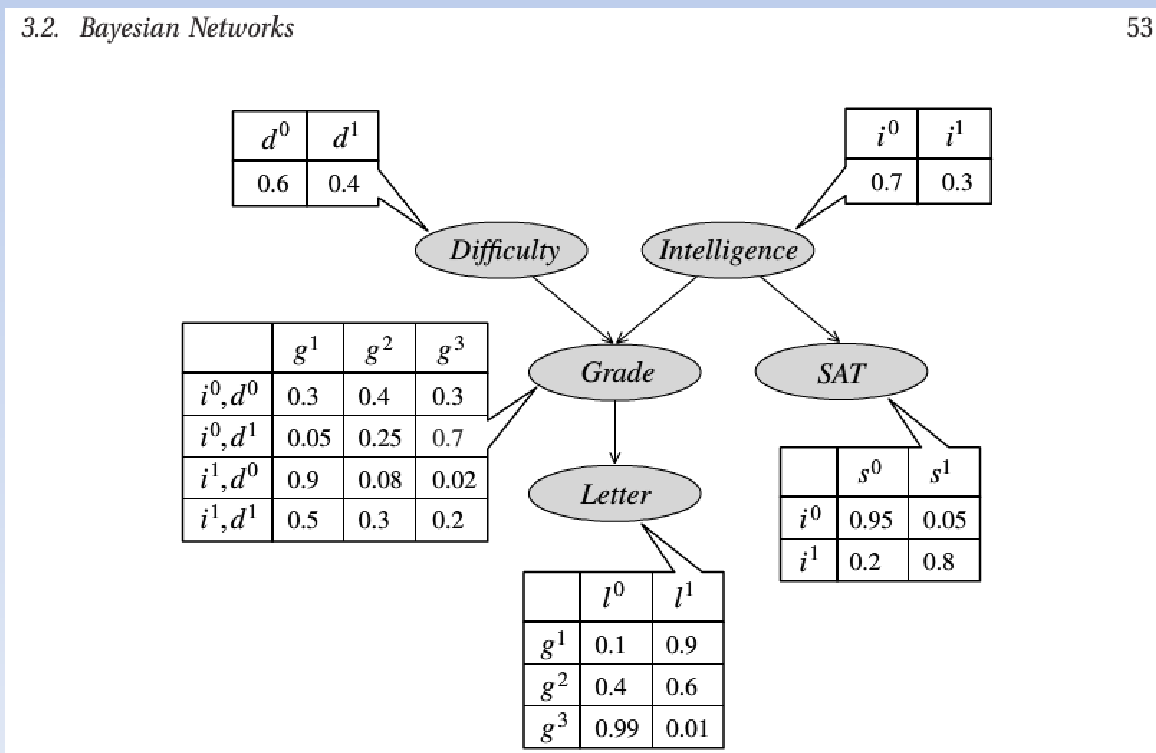
	g^1	g^2	g^3
i^0, d^0	0.3	0.4	0.3
i^0, d^1	0.05	0.25	0.7
i^1, d^0	0.9	0.08	0.02
i^1, d^1	0.5	0.3	0.2

	s^0	s^1
i^0	0.95	0.05
i^1	0.2	0.8

	l^0	l^1
g^1	0.1	0.9
g^2	0.4	0.6
g^3	0.99	0.01

Let’s Query w/ $\mathcal{B}^{\text{student}}$

- $P(i^1, d^0, g^2, s^1, l^0) =$
  - ➢ $P(i^1)P(d^0)P(g^2 | i^1, d^0)P(s^1 | i^1)P(l^0 | g^2)$
  - ➢ $0.3 \cdot 0.6 \cdot 0.08 \cdot 0.8 \cdot 0.4 =$



d^0	d^1
0.6	0.4

i^0	i^1
0.7	0.3

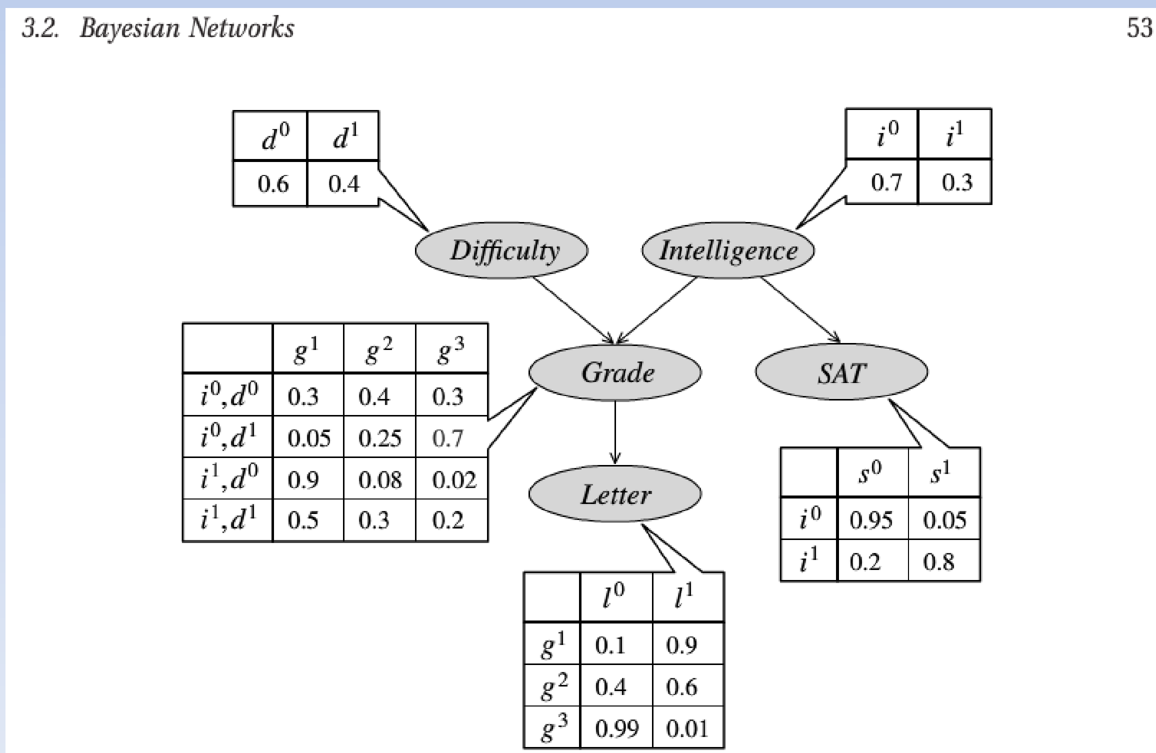
	g^1	g^2	g^3
i^0,d^0	0.3	0.4	0.3
i^0,d^1	0.05	0.25	0.7
i^1,d^0	0.9	0.08	0.02
i^1,d^1	0.5	0.3	0.2

	s^0	s^1
i^0	0.95	0.05
i^1	0.2	0.8

	l^0	l^1
g^1	0.1	0.9
g^2	0.4	0.6
g^3	0.99	0.01

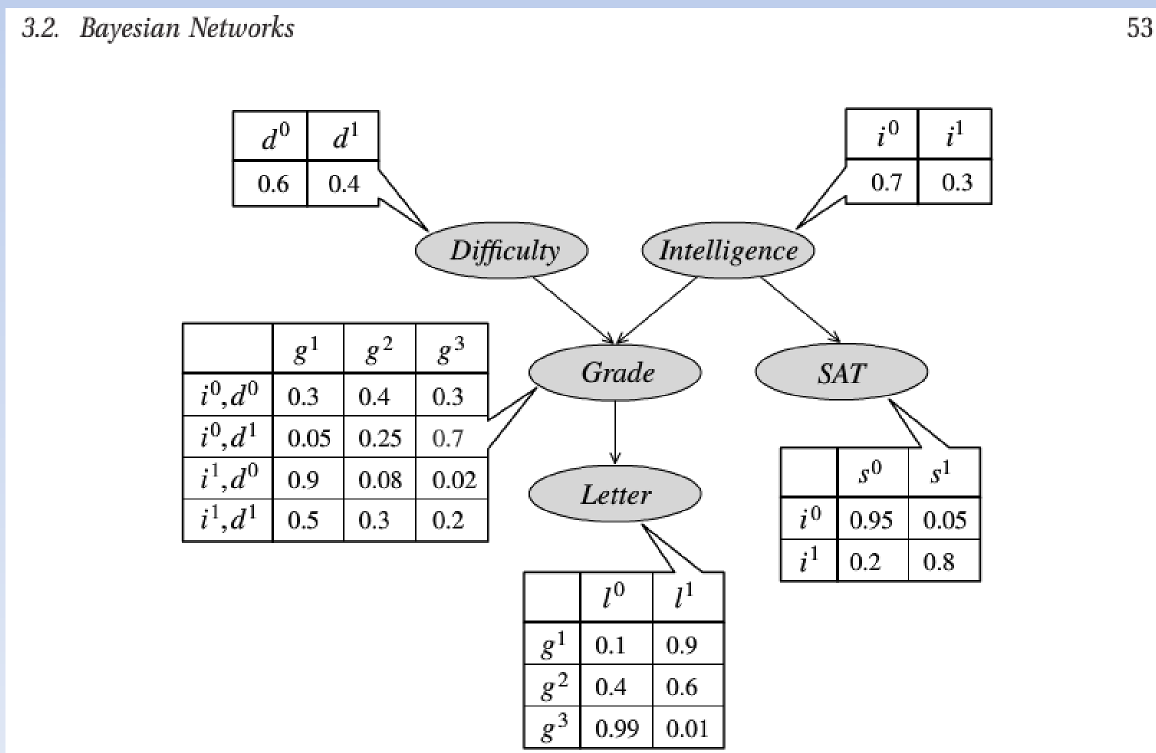
Let’s Query w/ $\mathcal{B}^{\text{student}}$

- $P(i^1, d^0, g^2, s^1, l^0)$ =
  - ➢ $P(i^1)P(d^0)P(g^2 | i^1, d^0)P(s^1 | i^1)P(l^0 | g^2)$
  - ➢ $0.3 \cdot 0.6 \cdot 0.08 \cdot 0.8 \cdot 0.4 = 0.004608$



Let's Query w/ $\mathcal{B}^{\text{student}}$

- $P(i^1, d^0, g^2, s^1, l^0) =$
  - ➢ $P(i^1)P(d^0)P(g^2 | i^1, d^0)P(s^1 | i^1)P(l^0 | g^2)$
  - ➢ $0.3 \cdot 0.6 \cdot 0.08 \cdot 0.8 \cdot 0.4 = 0.004608$????



d^0	d^1
0.6	0.4

i^0	i^1
0.7	0.3

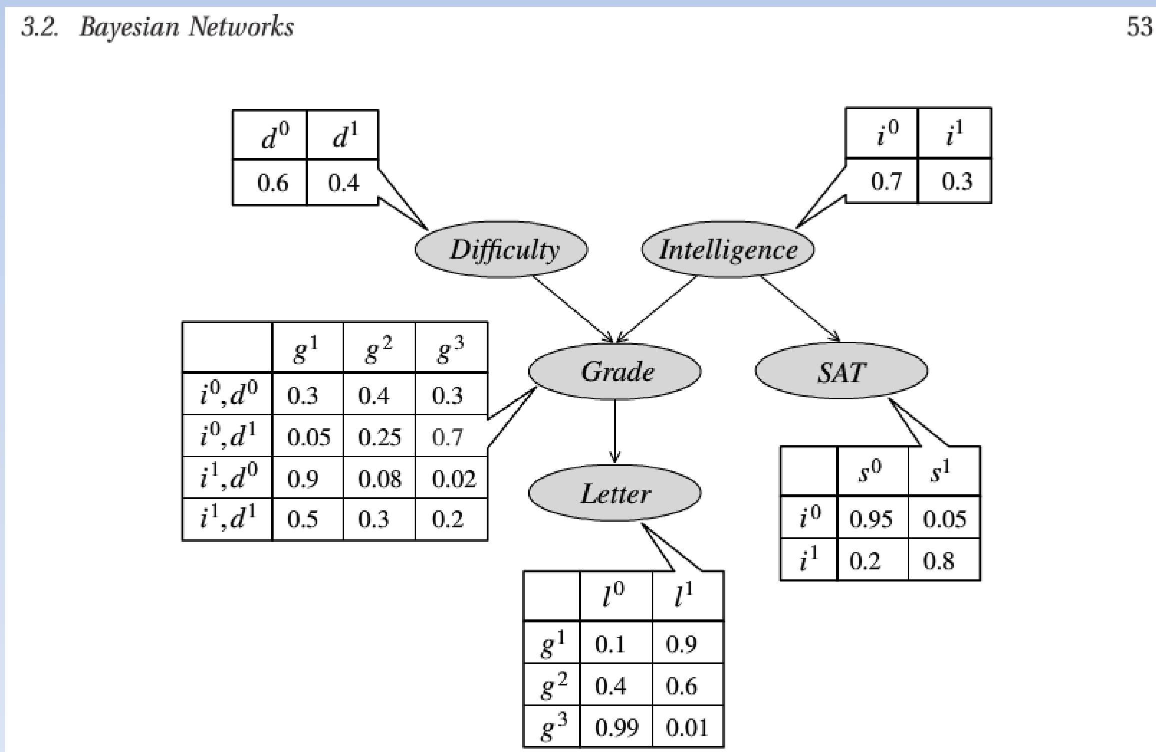
	g^1	g^2	g^3
i^0,d^0	0.3	0.4	0.3
i^0,d^1	0.05	0.25	0.7
i^1,d^0	0.9	0.08	0.02
i^1,d^1	0.5	0.3	0.2

	s^0	s^1
i^0	0.95	0.05
i^1	0.2	0.8

	l^0	l^1
g^1	0.1	0.9
g^2	0.4	0.6
g^3	0.99	0.01

First Example w/ Chain Rule for Bayesian Networks

- P(I, D, G, S, L)=
  - ➢P(I)P(D)P(G|I,D)P(S|I)P(L|G)



d^0	d^1
0.6	0.4

i^0	i^1
0.7	0.3

	g^1	g^2	g^3
i^0,d^0	0.3	0.4	0.3
i^0,d^1	0.05	0.25	0.7
i^1,d^0	0.9	0.08	0.02
i^1,d^1	0.5	0.3	0.2

	s^0	s^1
i^0	0.95	0.05
i^1	0.2	0.8

	l^0	l^1
g^1	0.1	0.9
g^2	0.4	0.6
g^3	0.99	0.01