Jd Joint Distribution

Uncertainty in AI: The Joint Distribution

CSE 415: Introduction to Artificial Intelligence
University of Washington
Winter, 2019

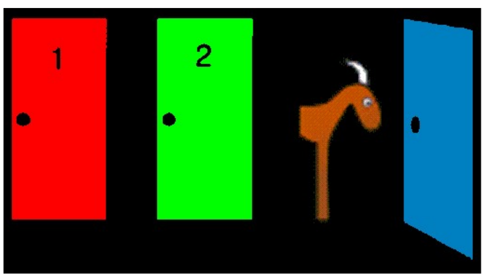
Jd Joint Distribution

Outline

- The Monty Hall Problem revisited
- Joint probability distributions
- Marginal distributions
- Factored joint probability distributions
- Bayes nets
- Benefits of Bayes nets for expert systems

Jd Joint Distribution

The Monty Hall Problem (revisited)



Jd Joint Distribution

Joint Probability Distribution for the Monty Hall Problem

Prize in	You choose	Host opens	P	Payoff if no switch	Payoff if switch
R	R	G	1/18	1	0
R	R	B	1/18	1	0
R	G	B	1/9	0	1
R	B	G	1/9	0	1
G	R	B	1/9	0	1
G	G	R	1/18	1	0
G	G	B	1/18	1	0
G	B	R	1/9	0	1
B	R	G	1/9	0	1
B	G	R	1/9	0	1
B	B	G	1/18	1	0
B	B	R	1/18	1	0

Jd Joint Distribution

Discussion

Marginal probability of winning, never switching: 1/3
Marginal probability of winning, always switching: 2/3

Other marginal probabilities:

P(prize is behind Red door) = 1/3
P(you choose Red door) = 1/3, assuming you choose randomly.
P(you first choose the right door) = 1/3

The joint probability distribution gives us the means to answer many questions about random variables and their relationships.

Jd Joint Distribution

Another Joint Distribution

Solar storm	Mike's battery on the fritz	Jan's radio works just fine	Mike's radio reception	Joint prob.
F	F	F	bad	0.0156
F	F	F	good	0.0468
F	F	F	none	0.0156
F	F	T	bad	0.1404
F	F	T	good	0.4212
F	F	T	none	0.1404
F	T	F	bad	0.0006
F	T	F	good	0.0002
F	T	F	none	0.0012
F	T	T	bad	0.0054
F	T	T	good	0.0018
F	T	T	none	0.0108

As our probability space gets larger, we need better techniques to handle our decision making

(continued)

Solar storm	Mike's battery on the fritz	Jan's radio works just fine	Mike's radio reception	Joint prob.
T	F	F	bad	0.0585
T	F	F	good	0.0117
T	F	F	none	0.0468
T	F	T	bad	0.039
T	F	T	good	0.0078
T	F	T	none	0.0312
T	T	F	bad	0.0006
T	T	F	good	0.00015
T	T	F	none	0.00225
T	T	T	bad	0.0004
T	T	T	good	0.0001
T	T	T	none	0.0015

Bayes Nets

Factored form of a joint distribution: uses less parameters in order to reduce our work

A practical way to manage probabilistic inference when multiple variables (perhaps many) are involved.

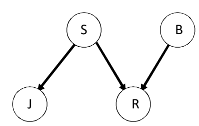
Requirement: The joint distribution is a "factored" distribution in which some random variables are either independent of or conditionally independent of most others.

Why Bayes Networks?

Reasoning about events involving many parts or contingencies generally requires that a joint probability distribution be known. Such a distribution might require thousands of parameters. Modeling at this level of detail is typically not practical.

Bayes Nets require making assumptions about the relevance of some conditions to others. Once the assumptions are made, the joint distribution can be "factored" so that there are many fewer separate parameters that must be specified.

Bayes Net for the Radio Problem



Directed acylic graph

Edge (u,v) tells us that u gives information about v.
i.e. v depends on u

S: [T, F] – "A solar storm is happening."
B: [T, F] – "Mike's battery is on the fritz."
J: [T, F] – "Jan's radio works just fine."
R: [bad, good, none] – "Mike's radio reception"

The variables of our system

variables dont have to be binary

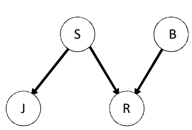
J is independent of B

If I know all parents

J and R are conditionally independent, both conditioned on S.

Conditionally independent: given that C happens, events A,B are conditionally independent given C iff knowing that A happened, given C gives us no information on whether or not B happened given C. Mathematically:
P(A and B I C) = P(A I C) P (B I C)

Factored Distribution



S and B have no parents; their distribution do not depend on any other factors

J depends on S as S is J's parent so its probability distribution is dependent on S

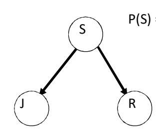
S: [P(T), P(F)] = [0.2, 0.8] sum up S = True probs in joint distribution
B: [P(T), P(F)] = [0.025, 0.975] sum up B = True probs in joint distribution
J: [P(T|S=T), P(F|S=T), P(T|S=F), P(F|S=F)] = [0.4, 0.6, 0.9, 0.1]
R: [P(bad|S=T,B=T), P(good|S=T,B=T), P(none|S=T,B=T), P(bad|S=T,B=F), …] = …

2*2*2*3 choices for parameter values in table

This factored distribution uses 20 parameters, rather than 24 for the unfactored version.

Not all of these are independent parameters: By using $\sum p_i = 1$, we can reduce the numbers to 12 and 23. For larger numbers of nodes, the savings are often much greater.

Working with the Bayes Net



P(S) = 0.2

P(J|S) = 0.4
P(J|~S) = 0.9

P(R=bad|S) = 0.493
P(R=good|S) = 0.099
P(R=none|S) = 0.409
P(R=bad|~S) = 0.203
P(R=good|~S) = 0.588
P(R=none|~S) = 0.210

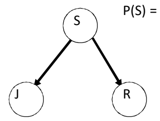
S: Solar Storm (A solar storm is happening.)
J: Jan's Radio (Jan's radio works just fine.)
R: Reception (Mike's radio's reception).

Suppose we know J & R (i.e. evidence), then we can infer properties about S using just Bayes Rule

For S, P(S = T) = 0.2 gives one free parameter, then P(F) is forced to be 1 - P(S = T)
Same for B
For J, P(J = T I S = T) gives one free parameter, then P(J = F I S = T) is forced to be 1 - P(J = T I S = T)
For J, P(J = T I S = F) gives one free parameter, then P(J = F I S = F) is forced to be 1 - P(J = T I S = F)

and there are 9 free parameters for R giving 12 total

Forward Propagation (from causes to effects)



P(S) = 0.2

P(J|S) = 0.4
P(J|~S) = 0.9

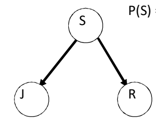
P(R=bad|S) = 0.493
P(R=good|S) = 0.099
P(R=none|S) = 0.409
P(R=bad|~S) = 0.203
P(R=good|~S) = 0.588
P(R=none|~S) = 0.210

Suppose S: there is a solar storm.
Then P(J|S) is 0.4, P(R=bad|S) = 0.493, etc.
Suppose ~S: no solar storm.
Then P(J|~S) is 0.9, P(R=bad|~S) = 0.203, etc.
(These come directly from the given information.)

Univ. of Wash. | The Joint Distribution | 13

going from parents probability distribution to children's probability distribution

Marginal Probabilities (using forward propagation)



P(S) = 0.2

P(J|S) = 0.4
P(J|~S) = 0.9

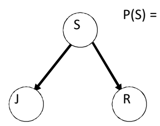
P(R=bad|S) = 0.493
P(R=good|S) = 0.099
P(R=none|S) = 0.409
P(R=bad|~S) = 0.203
P(R=good|~S) = 0.588
P(R=none|~S) = 0.210

Then P(J), the probability that J is true in any situation, is
P(J) = P(J|S)P(S) + P(J|~S)(1-P(S)) = 0.08 + 0.72 = 0.8
And P(R=bad), the prob. that R is bad in any situation, is
P(R=bad) = P(R=bad|S)P(S) + P(R=bad|~S)(1-P(S)) =
= (0.493)(0.2) + (0.203)(0.8) = 0.261

Marginalizing means eliminating a contingency by summing the probabilities for its different cases (here S and ~S).

Univ. of Wash. | The Joint Distribution | 14

Using Bayes' Rule



P(S) = 0.2

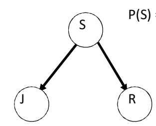
P(J|S) = 0.4
P(J|~S) = 0.9

P(R=bad|S) = 0.493
P(R=good|S) = 0.099
P(R=none|S) = 0.409
P(R=bad|~S) = 0.203
P(R=good|~S) = 0.588
P(R=none|~S) = 0.210

Suppose we know J: Jan's radio works just fine.
How do we update the probability of S?
Bayes' rule: P(S|J) = P(J|S)P(S)/P(J) = 0.08/0.8 = 0.1

Univ. of Wash. | The Joint Distribution | 15

Updating Probabilities of Consequences



P(S) = 0.2

P(J|S) = 0.4
P(J|~S) = 0.9

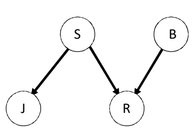
P(R=bad|S) = 0.493
P(R=good|S) = 0.099
P(R=none|S) = 0.409
P(R=bad|~S) = 0.203
P(R=good|~S) = 0.588
P(R=none|~S) = 0.210
P(R=bad|J) ?

Suppose we know Jan's radio works just fine.
How do we update the probability of R=bad?
Use the revised probability of S:
P(R=bad|J) = P(R=bad|S)P(S|J) + P(R=bad|~S)P(~S|J) =
(0.493)(0.1) + (0.203)(0.9) = 0.049 + 0.183 = 0.232
Which is slightly lower than P(R=bad) = 0.261.

Univ. of Wash. | The Joint Distribution | 16

Updating our prediction of the probability distribution of S.

Handling Multiple Causes



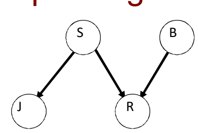
R	S	B	P(R|S,B)
R=bad	T	T	0.75
R=good	T	T	0.2
R=none	T	T	0.05
R=bad	T	F	0.4
R=good	T	F	0.5
R=none	T	F	0.1
R=bad	F	T	0.6
R=good	F	T	0.3
R=none	F	T	0.1
R=bad	F	F	0.2
R=good	F	F	0.2
R=none	F	F	0.6

B: [T, F] – "Mike's battery is on the fritz."

P(B) = 0.025

Univ. of Wash. | The Joint Distribution | 17

Explaining Away



R	S	B	P(R|S,B)
R=bad	T	T	0.75
R=good	T	T	0.2
R=none	T	T	0.05
R=bad	T	F	0.4
R=good	T	F	0.5
R=none	T	F	0.1
R=bad	F	T	0.6
R=good	F	T	0.3
R=none	F	T	0.1
R=bad	F	F	0.2
R=good	F	F	0.2
R=none	F	F	0.6

Suppose R=bad. This raises the probability for each cause:
P(S|R=bad) = 0.378, P(B|R=bad) = P(R=bad|B) P(B)/
P(R=bad) = 0.027

Now, in addition, suppose not J (Jan's radio not fine).
Not J makes it more likely that S is true,
"And this explains R=bad."
B is now less probable: P(B|R=bad, J=F) = 0.016.

Univ. of Wash. | The Joint Distribution | 18

Suppose we know information about J, i.e. that J is having issues with her radio. Then we might increase the probability of S, a solar storm occuring. This could explain why Mike is having an issue with his radio R, rather than an issue with his battery B. Thus, we might lower P(B = True). So we can see how information propagates through this network.

Benefits of Bayes Nets

"The decomposition of large probabilistic domains into weakly connected subsets through conditional independence is one of the most important developments in the recent history of AI."

(Russell & Norvig, 3e. p499.)

Benefits of Bayes Nets

The joint probability distribution with boolean random variables normally requires $2^n - 1$ independent parameters.

With Bayes Nets we only specify these parameters:

Less parameters

1. "root" node probabilities.
   e. g., P(A=true) = 0.2; P(A=false)=0.8.
2. For each non-root node, a table of 2^k values, where k is the number of parents of that node.
   Typically $k < 5$.
3. Propagating probabilities happens along the paths in the net. With a full joint prob. dist., many more computations may be needed.

information flows easier through the network as we see in slide 17