# Oct 14 Multivariate Probability

DF: drier, true

pairs. one-off

rows-> 4 3

5 ?

posterior

11 socks out of drier;

e.g. 4 pairs +3 one-off

P(# of socks we own GIVEN drier)

Socks wrapup. We had a model for pulling socks out of the drier after we'd simulated both how many socks we had and the underlying proportion of "paired" socks we owned. How was this going to be Bayes'?

$\underbrace{\text{P(drier w/ 11 socks GIVEN what we own)}}_{=A}$

1. It's conditional probability, not really Bayes'.
2. It's easy to model the drier outcomes GIVEN the true sock distribution, so Bayes' theorem lets us then use that conditional for the vice versa that we actually wanted: probabilities of sock distributions GIVEN what came out of the drier.
3. For events $s$ true socks, $d$ drier-removed-socks, $p$ true proportion, we have: $P(s,p|d) = \frac{P(d|s,p)P(s)P(p)}{P(d)}$ but we didn't really compute the denominator.
4. It kinda socked that we didn't fully finish that notebook on Friday.

Prior

$P(S|D) = \frac{P(S \text{ outcome AND drier outcome}) - \text{sims}}{P(\text{drier outcomes})}$ $\boxed{P(S|D)} > \frac{\overbrace{P(D|S)}^{A}\,\boxed{P(S)}}{P(D)}$

# Announcements and To-Dos

Announcements:

1. Working on Practicum official write-up:
   1.1 Implement and compare Simulated Annealing and Djikstra's algorithm on "The traveling salesman" problem.
   1.2 Short-ish paper on ethics and AI: think about a topic where AI might have important impacts on the world, get some sources, and write about the costs and benefits of certain AI tuning and implementation choices! Consider e.g. smartphones, GPS, game-playing, cars, USPS routes, national security, whatever interests you!

Last time we learned:

1. Bayes nets (and socks)

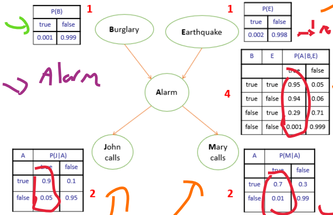
# Bayes Nets: the Alarm

Recall: Bayes nets encode joint distributions as the product of local conditionals:
$P(X_1 = x_1, X_2 = x_2, \ldots X_n = x_n) = \prod_{i=1}^{n} P(x_i | \text{parents}(X_i))$

cause: B, E

→ Alarm

Prob / each




→ $2^2 = 4$ total

→ 2 total

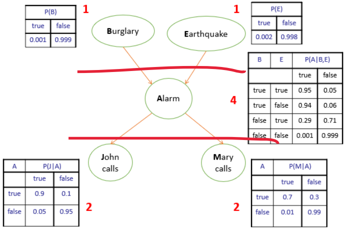
Calls from neighbors.

What is the entire joint distribution?

1. Events $B$ and $E$ are independent. So $P(B, E) = P(B)P(E)$. They're also class-conditionally independent *given* an alarm, but we don't even need that!
2. Event $A$ depends on $B$ and $E$. We know $P(A|B)$, $P(A|E)$, $P(A|BE)$. → specify 4 outcome
3. Once we know $A$, we know the class-independent probabilities of $P(J|A), P(M|A)$. Due to their independence, $P(JM|A) = P(J|A)P(M|A)$.

# Bayes Nets: the Alarm

Everything factors!




What is the entire joint distribution?

$P(B, E, A, J, M)$ → joint distribution

$= P(BE)P(AJM|BE)$ Defn. Conditional

$= P(B)P(E)P(AJM|BE)$ B & E ind.

$= P(B)P(E)P(A|BE)P(JM|ABE)$ Defn. Conditional

$= \underbrace{P(B)P(E)}_{\text{Priors}}\underbrace{P(A|BE)}_{\text{models}}\underbrace{P(J|A)P(M|A)}_{\text{model nbrs}}$ Class-cond. indep!

In effect we've divided the graph into tiers: a top tier of $B, E$, a middle tier of $A$, and the last tier of $J, M$. The overall probability is just the product of the statements on each tier conditioned on the tier above it!

# Bayes Nets: the Alarm

$$P(B, E, A, J, M) =$$

$$\boxed{P(B)P(E)P(A|BE)P(J|A)P(M|A)}$$

**Example:** What is
$P(B = True | J = True, M = True)$?

not asking about A & E: those "hidden"

$$P(B \mid J M) = \frac{P(B J M)}{P(JM)} = \frac{P(JM \mid B) \cdot P(B)}{P(JM)}$$

Bayes

Law total prob:

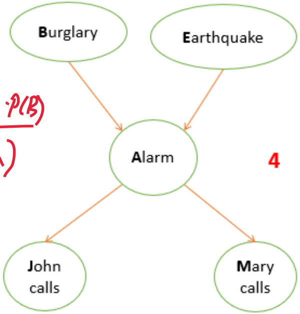
$P(A) = P(A \text{ AND } J=T, M=T)$
$+ P(A \text{ " } J=F, M=T)$
$+ P(A \text{ AND } J=T, M=F)$
$+ P(A \text{ AND } J=F, M=F)$

| P(B) | |
|---|---|
| true | false |
| 0.001 | 0.999 |

1

| P(E) | |
|---|---|
| true | false |
| 0.002 | 0.998 |

1




| B | E | P(A\|B,E) | |
|---|---|---|---|
| | | true | false |
| true | true | 0.95 | 0.05 |
| true | false | 0.94 | 0.06 |
| false | true | 0.29 | 0.71 |
| false | false | 0.001 | 0.999 |

4

need to include A & E

| A | P(J\|A) | |
|---|---|---|
| | true | false |
| true | 0.9 | 0.1 |
| false | 0.05 | 0.95 |

2

| A | P(M\|A) | |
|---|---|---|
| | true | false |
| true | 0.7 | 0.3 |
| false | 0.01 | 0.99 |

2

# Bayes Nets: the Alarm

$$P(B, E, A, J, M) =$$

$$P(B)P(E)P(A|BE)P(J|A)P(M|A)$$

**Example:** What is
$P(B = True|J = True, M = True)$?
**By hand Solution:**

$$P(B|JM) = \frac{P(BJM)}{P(JM)}$$

$P(BJM)$ happens 4 ways to add up: over both values of $E$ and both values of $A$. $P(JM)$ happens 8 ways to add up: over both values of $E$, both values of $A$, and both values of $B$.

That's 8 outcomes to consider! But they're all quick and easy! A specific outcome is on the tables. E.g.
$P(B = T, E = F, A = T, J = T, M = T)$
is one we need, and is
$P(B = T)P(E = F)P(A = T|B = T E = F)P(J = T|A = T)P(M = T|A = T)$ or
$(.001)(.998)(.94)(.9)(.7)$

| E | A | B |
|---|---|---|
| T | T | |
| F | T | |
| T | F | |
| F | F | |

1

| P(B) | |
|---|---|
| true | false |
| 0.001 | 0.999 |

1

| P(E) | |
|---|---|
| true | false |
| 0.002 | 0.998 |


4

| B | E | P(A\|B,E) | |
|---|---|---|---|
| | | true | false |
| true | true | 0.95 | 0.05 |
| true | false | 0.94 | 0.06 |
| false | true | 0.29 | 0.71 |
| false | false | 0.001 | 0.999 |

2

| A | P(J\|A) | |
|---|---|---|
| | true | false |
| true | 0.9 | 0.1 |
| false | 0.05 | 0.95 |

2

| A | P(M\|A) | |
|---|---|---|
| | true | false |
| true | 0.7 | 0.3 |
| false | 0.01 | 0.99 |

$P(\text{all } 5) = P(B)\,P(E) \cdot P(A|BE) \cdot P(M|A) \cdot P(J|A)$

# Bayes Nets: the Alarm

$$P(B, E, A, J, M) =$$

$$P(B)P(E)P(A|BE)P(J|A)P(M|A)$$

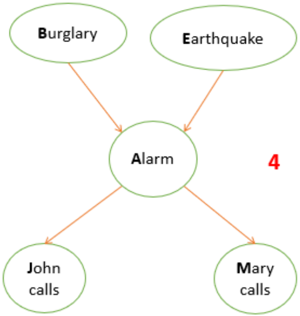
**Example:** What is
$P(B = True|J = True, M = True)$?

1

| P(B) | |
|---|---|
| true | false |
| 0.001 | 0.999 |

1

| P(E) | |
|---|---|
| true | false |
| 0.002 | 0.998 |




4

| B | E | P(A\|B,E) | |
|---|---|---|---|
| | | true | false |
| true | true | 0.95 | 0.05 |
| true | false | 0.94 | 0.06 |
| false | true | 0.29 | 0.71 |
| false | false | 0.001 | 0.999 |

2

| A | P(J\|A) | |
|---|---|---|
| | true | false |
| true | 0.9 | 0.1 |
| false | 0.05 | 0.95 |

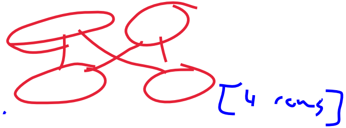
2

| A | P(M\|A) | |
|---|---|---|
| | true | false |
| true | 0.7 | 0.3 |
| false | 0.01 | 0.99 |

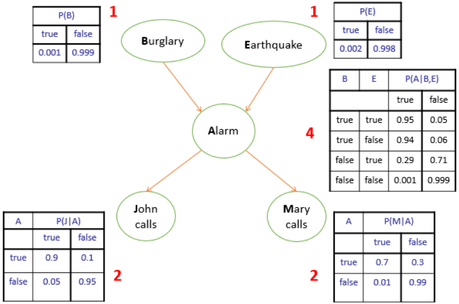
That's 8 outcomes to consider! But they're all quick and easy! A specific outcome is on the tables. E.g. $P(B = T, E = F, A = T, J = T, M = T)$ is one we need, and is $P(B = T)P(E = F)P(A = T|B = TE = F)P(J = T|A = T)P(M = T|A = T)$ or $(.001)(.998)(.94)(.9)(.7)$

# Bayes Nets: Space

Are we really saving space and time, here?

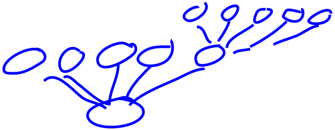



The full joint distribution on $n = 5$ nodes would take $2^5 = 32$ probabilities. We're specifying only 10.




Suppose the Bayes net in general has $k = 2$ parents per node. Then for $n = 5$ nodes we're only specifying $\cancel{25 \cdot 2^2} = 20$ probabilities at worst!

$5 \cdot 2^2$



But what if $n = 30$ and $k = 5$? The Bayes net would require $n \cdot 2^k = 960$ probabilities. The full joint distribution holds $2^{30} \approx 1e9$ entries.

# Bayes Nets: Enumeration

It turns out there was a faster way to represent the desired calculation of

$$P(B|JM) = \frac{P(BJM)}{P(JM)}.$$

→ summing 4 exact outcomes, $(A, E)^2$

→ summing 8 exact outcomes $(A, E, B)^2$

First, let $\alpha = \frac{1}{P(JM)}$. This is the entire denominator, and it's actually pretty unimportant! We then write

It turns out there was a faster way to represent the desired calculation of

$$P(B|JM) = \alpha P(BJM).$$

We're going to add the events for $A$ and $E$ back into the calculation, since those are the things we have and need!

## Enumeration, continued

AND = joint distribution

$$P(B|JM) = \frac{P(BJM)}{P(JM)}$$ ← just conditioning — Def'n conditional prob

$$= \alpha P(BJM)$$ — ignore denom for now

$$= \alpha \sum_A P(BJM|A)P(A)$$ — marginal over A

Law total prob

each possible level of A

$$= \alpha \sum_A P(BJMA)$$ — joint with A

AND

$$= \alpha \sum_A \sum_E P(BJMA|E)P(E)$$ — marginal over A

L.T.P.

include E

$$= \alpha \sum_E \sum_A P(BJMAE)$$ — joint with E

→ reverse order

ex. plug in

B = True
J = True
M = True

| M | J | P(B=T) | P(B=F) |
|---|---|---|---|
| T | T | a | 1-a |
| T | F | b | 1-b |
| F | T | c | 1-c |
| F | F | d | 1-d |

| E | A | P(BJMAE) |
|---|---|---|
| T | T | |
| T | F | |
| F | T | |
| F | F | |

# Enumeration, payoff

P(B & J & m)

All told we have:

$$P(B|JM) = \alpha \sum_E \sum_A P(BJMAE)$$

denom

But the Bayes' net tells us that the full joint distribution can be broken into the conditional-on-parents parts:

if B & E know A Prob

$$P(B|JM) = \alpha \sum_E \sum_A \prod_{i=1}^{n} P(X_i|\text{parents}(X_i))$$

if A know J & m

which for us is:

$$P(B|JM) = \alpha \sum_E \sum_A P(B) P(E) P(A|BE) P(J|A) P(M|A)$$

$$= \alpha P(B) \sum_E P(E) \sum_A P(A|BE) P(J|A) P(M|A)$$

B back in E back in P on A

# The Denominator

Part of why we skip the denominator on a discrete outcome is it comes from the prior calculations. Suppose we use enumeration, and find that e.g.

$P(B=F \mid JM)$
$P(B=True \mid JM)$

$=T/F$ $\quad$ $=T/F$

$$P(B|JM) = \alpha \sum_E \sum_A P(B)P(E)P(A|BE)P(J|A)P(M|A)$$

$$P(+B|JM) = \alpha .07 \qquad = \alpha \cdot P(B \text{ and } JM)$$

$$P(\neg B|JM) = \alpha .02 \qquad = \alpha \cdot P(\text{not } B \text{ and } JM)$$

Then we have $\alpha$. We know that $P(+B|JM) + P(-B|JM) = 1$, so $\alpha .07 + \alpha .02 = 1 \implies \alpha = \frac{1}{.09}$. We can reuse our work in the numerator and compute the complementing outcome... possibly as we go!

$1/(.07+.02)$

# Bayes Nets: Vocab

$P(X_1 \,\&\, X_2) = P(X_1) \cdot P(X_2 | X_1)$

parent

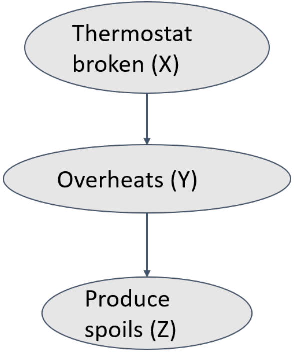
When describing Bayes' nets, we often try to ensure that we're maintaining a proper order of nodes. The goal is to list parents as causes before their effects. If we don't, that's kind of OK, but we'd have to specify the conditional probabilities backwards: we'd end up with e.g. $P(rain|sidewalks\, wet)$.

causal

**Definition:** A ~~*casual*~~ *chain* is shown to the right.

1. Are $X$ and $Z$ independent? NO
2. Knowledge of $Z$ should influence or *propagate* belief about $X$.

What about $X$ and $Z$ *given* $Y$?

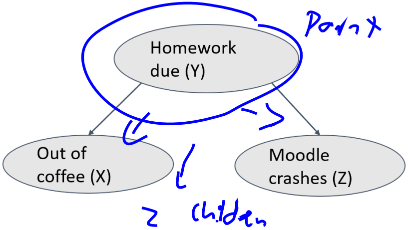



- lowest levels are often observed

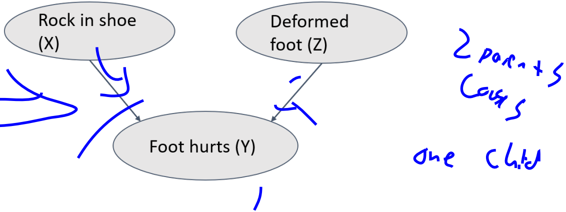
- infer higher up

or we ask hypotheticals: how does an increased P(E) affect M calling?

# Bayes Nets: Common Causes and Effects




Swap!




**Definition:** A *common cause* is a children with two parents in the network. Multiple events could be attributed to causing the observed node.

**Definition:** A *common effect* is a parent with two children in the network. One event causes multiple effects.

Are $X$ and $Z$ independent? Are they independent *given* $Y$?

# Bayes Nets: Infrastructure

variables edges

What do we need to describe a Bayesian Network? Nodes, Arcs, Conditional Probability Tables (CPTs).

particular_bayes_net = BayesNet([list of nodes: ('Name', 'Parents', [T, F], dict cpt)])

A | B

A | B, C

A | "parents"

class BayesNet:

- generic "tree class" that will interact with a more specific Node class
- Read in the "Node Specifications," like what's listed above.
- Add Nodes using class BayesNode

class BayesNode:

- node = BayesNode(name=name, parents = parents, values = values, cpt = cpt)

And some functions:

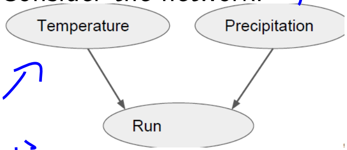
- We'll have a probability function - calculates the probability of seeing a particular BayesNode variable when the parents' values are given as "evidence" (could be on table: np.random.choice .uniform)
- We'll have a class PDF_discrete
- Lastly, we'll need a way to compute probabilities and "ask" things of our Bayesian Network: enumeration_ask, enumeration_all

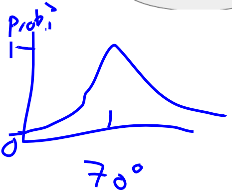
# Continuity

We may want to have a *continuous* parent.

Consider the network:

T/F




$P(R|T)$

# Continuity

T, Rain : R+

P(T, Rain): R 0,1

Temp

We may want to have a *continuous* parent.

- One way is to specify a function that gives a single probability of running *given* parents. We really need *any* function that takes in continuous numbers and outputs a probability.

  One option, *logit*:

$$P(x) = \frac{1}{1+e^{-x}}$$

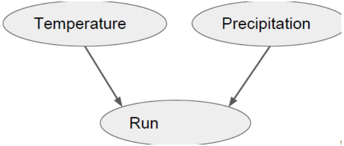
$e^{-x} \in (0, \infty)$

70° ?

  Another option, *probit*:

$$P(x) = \int_{-\infty}^{x} N(0,1)(y)\,dy$$

Consider the network:




input: Temp

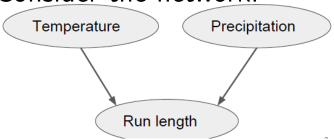
output: [0,1]: P(R|T)

- Alternatively, discretize: specify probabilities of going for a run for specific *ranged* of temperature/precipitation. The cutoffs we might use are called *thresholds*.

# More Continuity

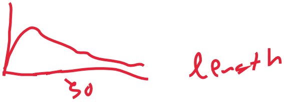
What if the response is continuous, itself?

Consider the network:




pdf!

Given T=60°, Rain=False



# More Continuity

Make $\mu$ a function of $T$, $R$.

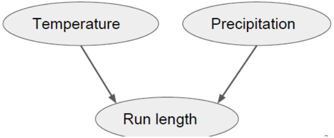
What if the response is continuous, itself?

1. Sums could just be integrals, where we specify a continuous pdf whose parameters move with the discrete predictors.
   - **Example:**

$$f_{length}(x) \sim \begin{cases} N(\mu_r(T), \sigma) & \text{if raining} \\ N(\mu_{-r}(T), \sigma) & \text{if not raining} \end{cases}$$

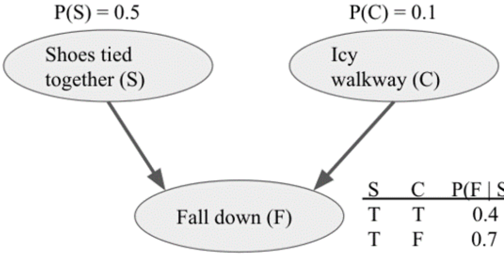
2. We could discretize: treat ranges and binary outcomes instead and answer e.g. P(run length<50).

Consider the network:



# Sampling Example

Consider the network:




| S | C | P(F \| S, C) |
|---|---|---|
| T | T | 0.4 |
| T | F | 0.7 |
| F | T | 0.1 |
| F | F | 0.01 |

To *sample* on a Bayesian network, we move from the top down.

1. Sample from the priors. These are nodes without parents.
2. Sample from the conditionals. These are children given their parents.
3. Save a data frame with all the outcomes.

The data frame is holding *joint* outcomes, so whatever conditional probabilities we desire live in it's rows and just counting outcomes with Booleans!

# Sampling Example

To *sample* on a Bayesian network, we move from the top down.

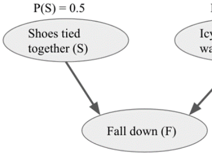
1. Sample from the priors. These are nodes without parents.
2. Sample from the conditionals. These are children given their parents.
3. Save a data frame with all the outcomes.

We may have a DFSAMPLES:

| Sample # | S | C | F |
|---|---|---|---|
| 0 | T | T | T |
| 1 | T | T | F |
| 2 | T | F | T |
| 3 | F | T | F |
| 4 | T | T | F |
| 5 | F | T | F |
| 6 | T | T | T |
| 7 | F | F | F |
| ⋮ | ⋮ | ⋮ | ⋮ |

We can check $P(C = T|F = T) = \frac{P(C=T \textbf{ AND } F=T)}{P(F=T)}$ by counting the numbers of rows where each set of Booleans is true.

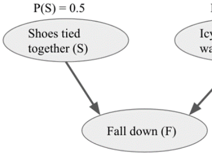
# Enumeration Example




| S | C | P(F \| S, C) |
|---|---|---|
| T | T | 0.4 |
| T | F | 0.7 |
| F | T | 0.1 |
| F | F | 0.01 |

What is the fully enumerated probability $P(+S|-F)$?

# Enumeration Example




| S | C | P(F \| S, C) |
|---|---|---|
| T | T | 0.4 |
| T | F | 0.7 |
| F | T | 0.1 |
| F | F | 0.01 |

What is the fully enumerated probability $P(+S|-F)$?

$$P(S|F) = P(S \textbf{ AND } F)/P(F)$$
$$P(S|F) = \alpha P(S)P(F|S)$$
$$P(S|F) = \alpha P(S)\left[P(F|(S \textbf{ AND } +C) + P(F|(S \textbf{ AND } \neg C)\right]$$
$$P(S|F) = \alpha P(S)\sum_I P(F|SC)P(C)$$
$$P(+S|\neg F) = \alpha P(+S)\sum_C P(\neg F|+S,C)P(C)$$
$$P(+S|\neg F) = \alpha .5\,(.1 \cdot .4 + .9 \cdot .7)$$
$$P(-S|\neg F) = \alpha P(-S)\sum_C P(\neg F|-S,C)P(C)$$
$$P(+S|\neg F) = \alpha .5\,(.1 \cdot .1 + .9 \cdot .01)$$

# Moving Forward

- Coming up:
    1. Markov!