Bayesian Networks

You will be expected to know

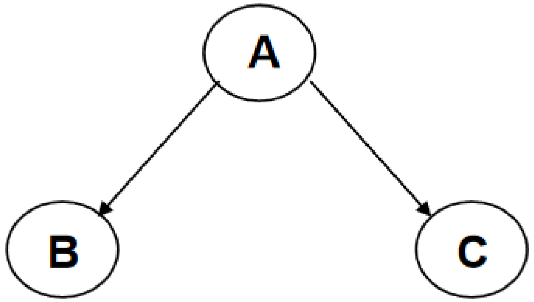
- Basic concepts and vocabulary of Bayesian networks.
  - Nodes represent random variables.
  - Directed arcs represent (informally) direct influences.
  - Conditional probability tables, P(Xi | Parents(Xi)).

- Given a Bayesian network:
  - Write down the full joint distribution it represents.

- Given a full joint distribution in factored form:
  - Draw the Bayesian network that represents it.

- Given a variable ordering and some background assertions of conditional independence among the variables:
  - Write down the factored form of the full joint distribution, as simplified by the conditional independence assertions.

Extended example of 3-way Bayesian Networks



Conditionally independent effects:
$p(A,B,C) = p(B|A)p(C|A)p(A)$

B and C are conditionally independent
Given A

E.g., A is a disease, and we model
B and C as conditionally independent
symptoms given A

E.g., A is Fire, B is Heat, C is Smoke.
“Where there’s Smoke, there’s Fire.”

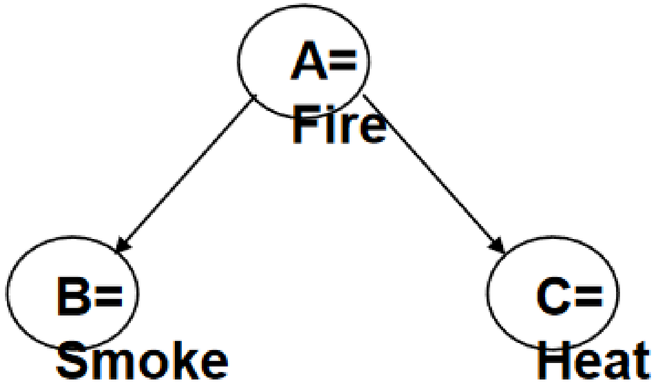
If we see Smoke, we can infer Fire.

If we see Smoke, observing Heat tells
us very little additional information.

Extended example of 3-way Bayesian Networks

Suppose I build a fire in my fireplace about once every 10 days…

P(fire)
0.1



Fire	P(Smoke)
t	.90
f	.001

Fire	P(Heat)
t	.99
f	.0001

Conditionally independent effects:
$P(A,B,C) = P(B|A)P(C|A)P(A)$

Smoke and Heat are conditionally independent given Fire.

If we see B=Smoke, observing C=Heat tells us very little additional information.

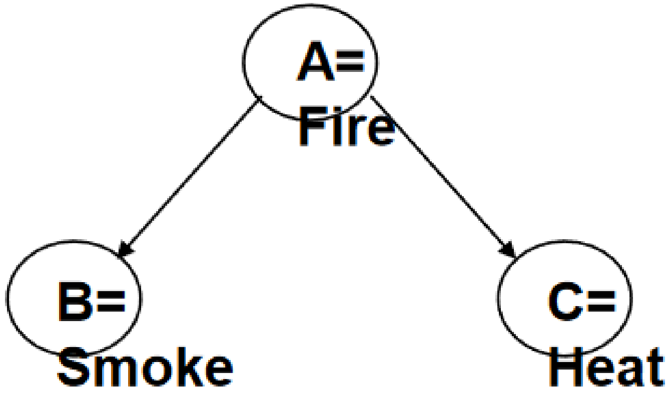
Extended example of 3-way Bayesian Networks

What is P(Fire=t | Smoke=t)?

P(Fire=t | Smoke=t)

=P(Fire=t & Smoke=t) / P(Smoke=t)

P(Fire)
0. 1



Fire	P(Smoke)
t	.90
f	.001

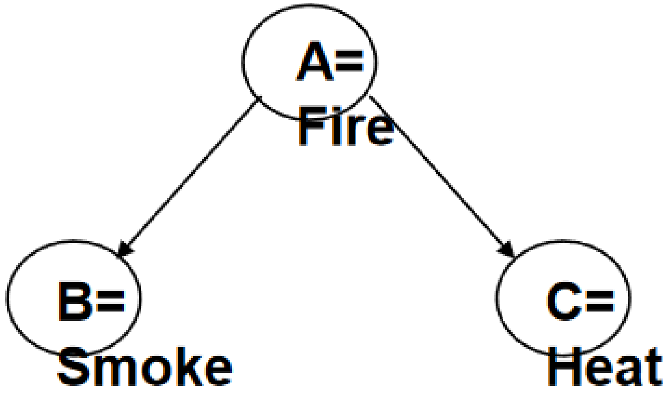
Fire	P(Heat)
t	.99
f	.0001

Extended example of 3-way Bayesian Networks

What is P(Fire=t & Smoke=t)?

P(Fire=t & Smoke=t)
=Σ_heat P(Fire=t&Smoke=t&heat)
=Σ_heat P(Smoke=t&heat|Fire=t)P(Fire=t)
=Σ_heat P(Smoke=t|Fire=t) P(heat|Fire=t)P(Fire=t)
=P(Smoke=t|Fire=t) P(heat=t|Fire=t)P(Fire=t)
+P(Smoke=t|Fire=t)P(heat=f|Fire=t)P(Fire=t)
= (.90x.99x.1)+(.90x.01x.1)
= 0.09

P(Fire)
0. 1

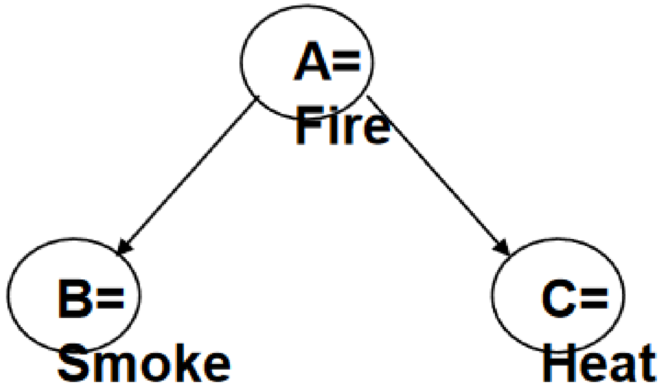


Fire	P(Smoke)
t	.90
f	.001

Fire	P(Heat)
t	.99
f	.0001

Extended example of 3-way Bayesian Networks

P(Fire)
0. 1



Fire	P(Smoke)
t	.90
f	.001

Fire	P(Heat)
t	.99
f	.0001

What is P(Smoke=t)?

P(Smoke=t)

$=\Sigma\_fire\ \Sigma\_heat\ P(Smoke=t\&fire\&heat)$

$=\Sigma\_fire\ \Sigma\_heat\ P(Smoke=t\&heat|fire)P(fire)$

$=\Sigma\_fire\ \Sigma\_heat\ P(Smoke=t|fire)\ P(heat|fire)P(fire)$

$=P(Smoke=t|fire=t)\ P(heat=t|fire=t)P(fire=t)$

$+P(Smoke=t|fire=t)P(heat=f|fire=t)P(fire=t)$

$+P(Smoke=t|fire=f)\ P(heat=t|fire=f)P(fire=f)$

$+P(Smoke=t|fire=f)P(heat=f|fire=f)P(fire=f)$

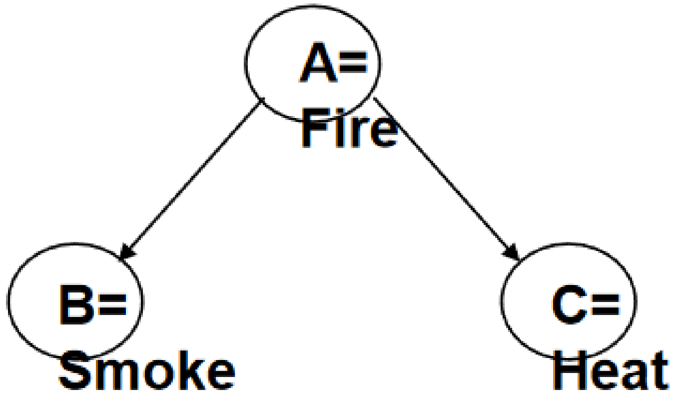
$= (.90x.99x.1)+(.90x.01x.1)$

$+(.001x.0001x.9)+(.001x.9999x.9)$

≈ 0.0909

Extended example of 3-way Bayesian Networks

P(Fire)
0. 1



Fire	P(Smoke)
t	.90
f	.001

Fire	P(Heat)
t	.99
f	.0001

What is P(Fire=t | Smoke=t)?

P(Fire=t | Smoke=t)

=P(Fire=t & Smoke=t) / P(Smoke=t)

$\approx$ 0.09 / 0.0909

$\approx$ 0.99

So we've just proven that

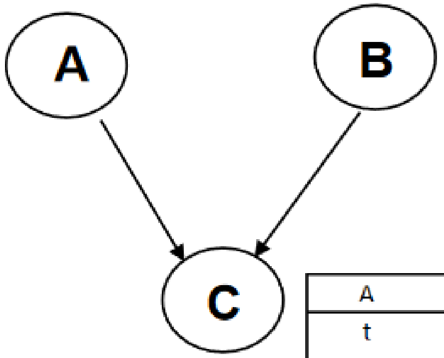
"Where there's smoke, there's (probably) fire."

Bayesian Network

- A Bayesian network specifies a joint distribution in a structured form:

$$p(A,B,C) = p(C|A,B)p(A)p(B) \quad \longleftrightarrow$$

P(A)
0.33



P(B)
0.67

A	B	P(C)
t	t	0.2
t	f	0.4
f	t	0.3
f	f	0.3

- Dependence/independence represented via a directed graph:
  - Node = random variable
  - Directed Edge = conditional dependence
  - Absence of Edge = conditional independence

- Allows concise view of joint distribution relationships:
  - Graph nodes and edges show conditional relationships between variables.
  - Tables provide probability data.

Bayesian Networks

- Structure of the graph ⇔ Conditional independence relations

  In general,

  $$p(X_1, X_2, \ldots X_N) = \prod p(X_i \mid parents(X_i))$$

  The full joint distribution

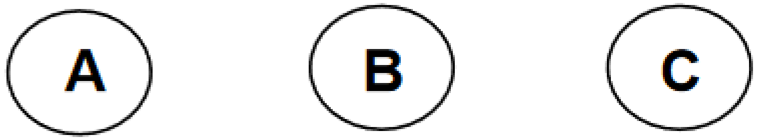
  The graph-structured approximation

- Requires that graph is acyclic (no directed cycles)

- 2 components to a Bayesian network
  - The graph structure (conditional independence assumptions)
  - The numerical probabilities (for each variable given its parents)

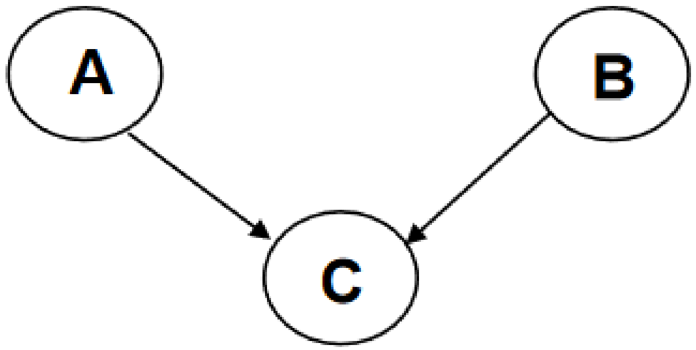
- Also known as belief networks, graphical models, causal networks

Examples of 3-way Bayesian Networks



Marginal Independence:
$p(A,B,C) = p(A)\, p(B)\, p(C)$

Examples of 3-way Bayesian Networks

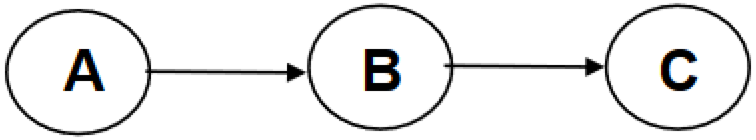


Independent Causes:
p(A,B,C) = p(C|A,B)p(A)p(B)

“Explaining away” effect:
Given C, observing A makes B less likely
e.g., earthquake/burglary/alarm example

A and B are (marginally) independent
but become dependent once C is known

Examples of 3-way Bayesian Networks



Markov dependence:
p(A,B,C) = p(C|B) p(B|A)p(A)

Burglar Alarm Example

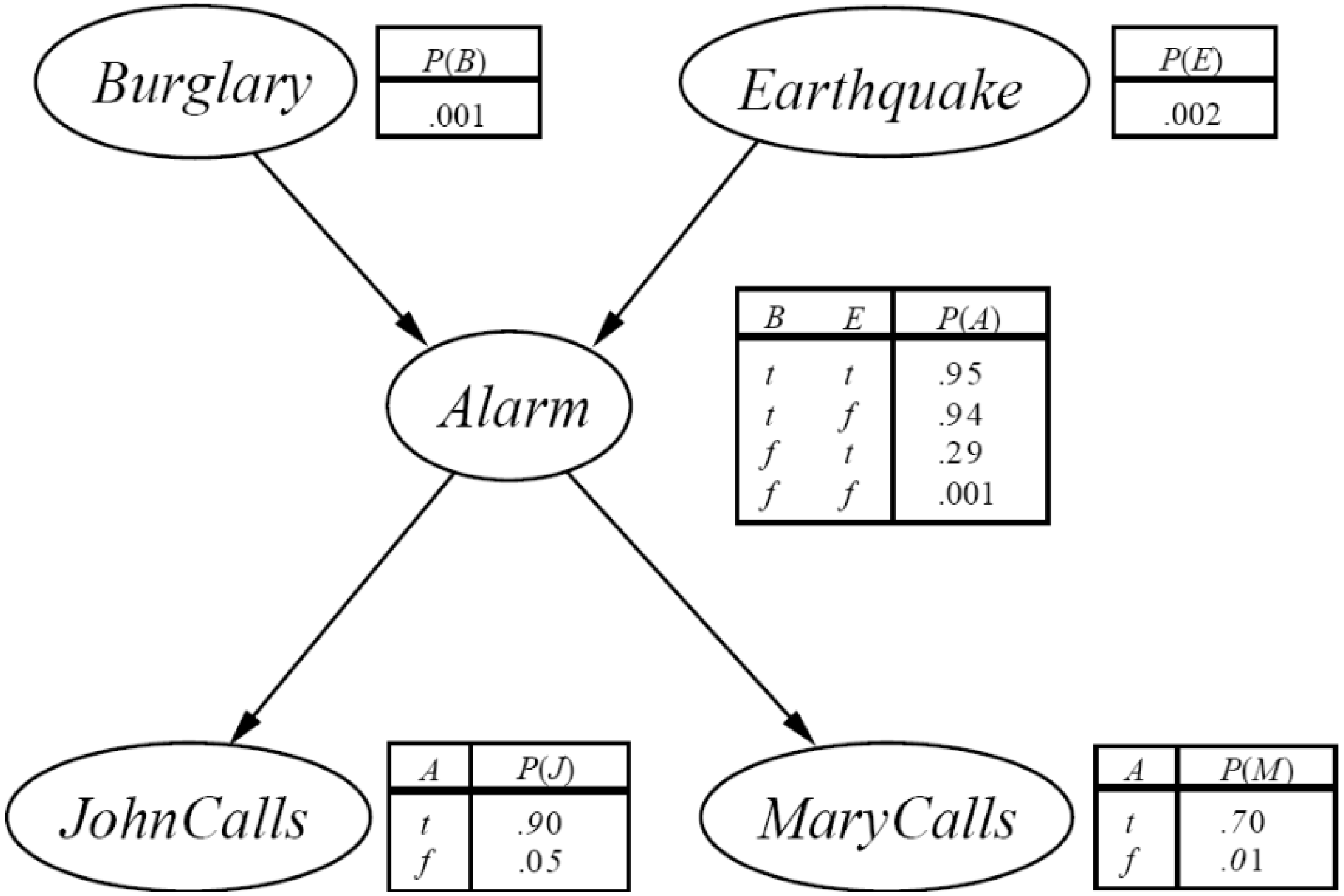
- Consider the following 5 binary variables:
  - B = a burglary occurs at your house
  - E = an earthquake occurs at your house
  - A = the alarm goes off
  - J = John calls to report the alarm
  - M = Mary calls to report the alarm

  - What is P(B | M, J) ? (for example)

  - We can use the full joint distribution to answer this question
    - Requires $2^5 = 32$ probabilities

    - Can we use prior domain knowledge to come up with a Bayesian network that requires fewer probabilities?

The Desired Bayesian Network



P(B)
.001

P(E)
.002

B	E	P(A)
t	t	.95
t	f	.94
f	t	.29
f	f	.001

A	P(J)
t	.90
f	.05

A	P(M)
t	.70
f	.01

Only requires 10 probabilities!

Constructing a Bayesian Network: Step 1

- Order the variables in terms of influence (may be a partial order)

  e.g., {E, B} -> {A} -> {J, M}

- $P(J, M, A, E, B) = P(J, M \mid A, E, B)\, P(A \mid E, B)\, P(E, B)$

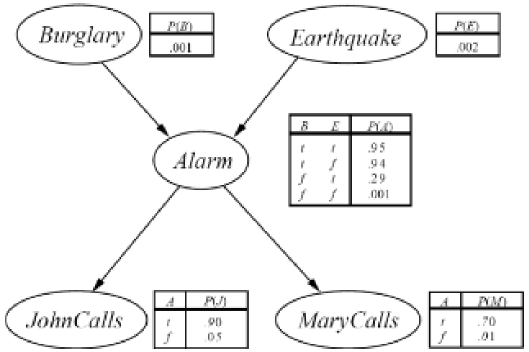
$$\approx P(J, M \mid A) \qquad P(A \mid E, B)\, P(E)\, P(B)$$

$$\approx P(J \mid A)\, P(M \mid A)\, P(A \mid E, B)\, P(E)\, P(B)$$

These conditional independence assumptions are reflected in the graph structure of the Bayesian network

Constructing this Bayesian Network: Step 2

- P(J, M, A, E, B) =
  P(J | A) P(M | A) P(A | E, B) P(E) P(B)

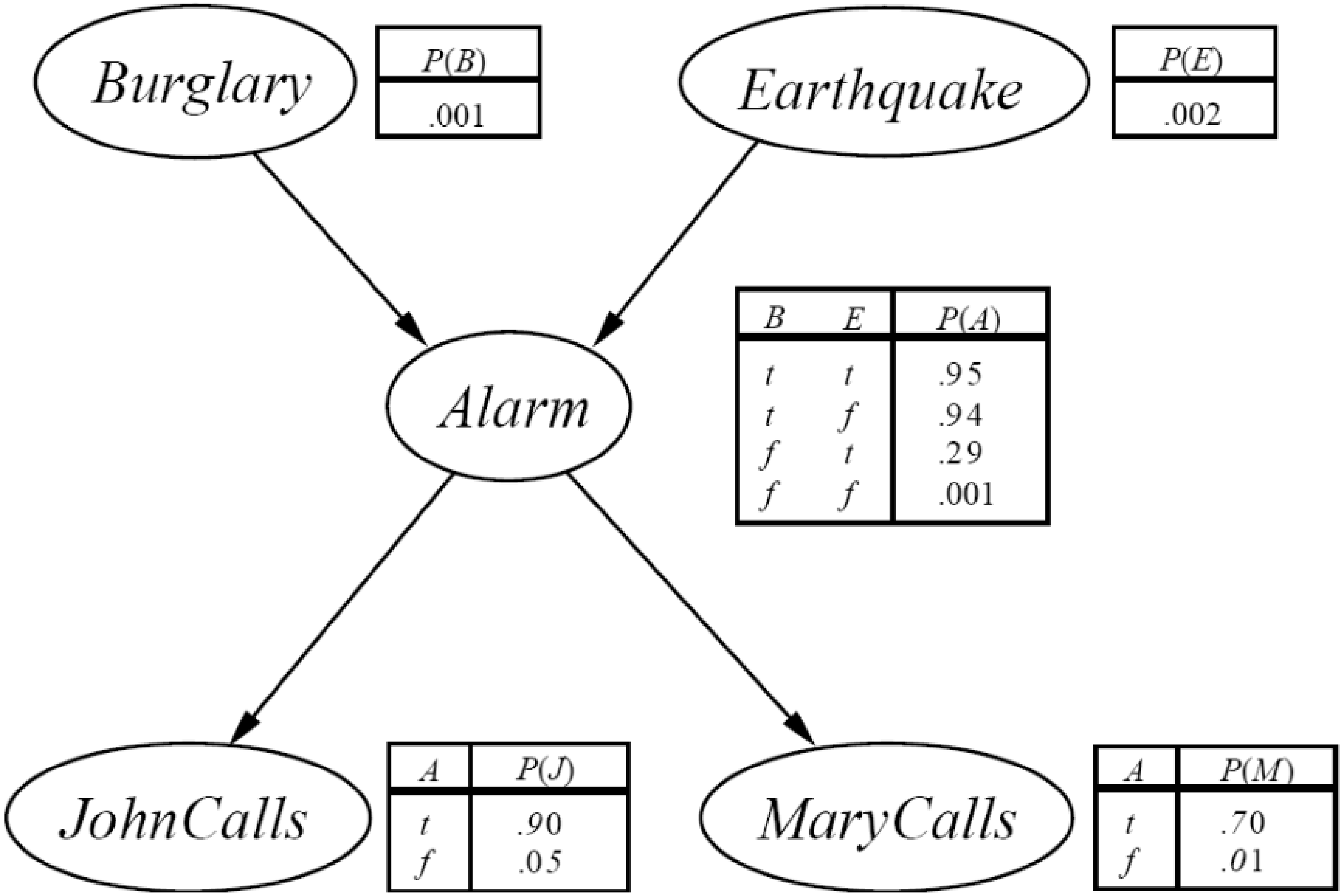


- There are 3 conditional probability tables (CPDs) to be determined:
  P(J | A), P(M | A), P(A | E, B)
  - Requiring 2 + 2 + 4 = 8 probabilities

- And 2 marginal probabilities P(E), P(B) -> 2 more probabilities

- Where do these probabilities come from?
  - Expert knowledge
  - From data (relative frequency estimates)
  - Or a combination of both - see discussion in Section 20.1 and 20.2 (optional)

The Resulting Bayesian Network



P(B)
.001

P(E)
.002

B	E	P(A)
t	t	.95
t	f	.94
f	t	.29
f	f	.001

A	P(J)
t	.90
f	.05

A	P(M)
t	.70
f	.01

Example of Answering a Probability Query

- So, what is $P(B \mid M, J)$?
  E.g., say, $P(b \mid m, \neg j)$, i.e., $P(B{=}true \mid M{=}true \wedge J{=}false)$

$P(b \mid m, \neg j) = P(b, m, \neg j) / P(m, \neg j)$;by definition

$P(b, m, \neg j) = \Sigma A \in \{a, \neg a\} \Sigma E \in \{e, \neg e\}\ P(\neg j, m, A, E, b)$;marginal

$P(J, M, A, E, B) \approx\ P(J \mid A)\ P(M \mid A)\ P(A \mid E, B)\ P(E)\ P(B)$; conditional indep.
$P(\neg j, m, A, E, b) \approx\ P(\neg j \mid A)\ P(m \mid A)\ P(A \mid E, b)\ P(E)\ P(b)$

Say, work the case $A{=}a \wedge E{=}\neg e$
$P(\neg j, m, a, \neg e, b) \approx\ P(\neg j \mid a)\ P(m \mid a)\ P(a \mid \neg e, b)\ P(\neg e)\ P(b)$
$\approx\ 0.10\ \times\ 0.70\ \times\ 0.94\ \times\ 0.998 \times 0.001$
Similar for the cases of $a \wedge e$, $\neg a \wedge e$, $\neg a \wedge \neg e$.

Similar for $P(m, \neg j)$. Then just divide to get $P(b \mid m, \neg j)$.