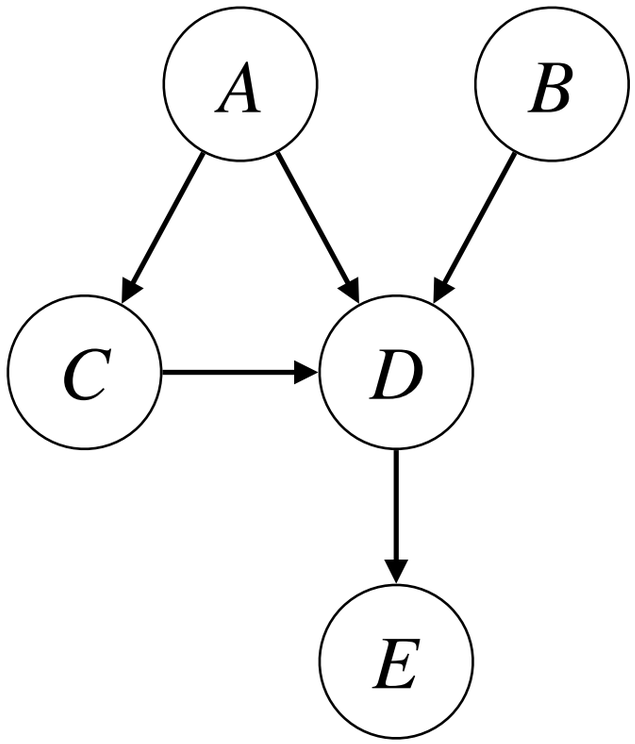
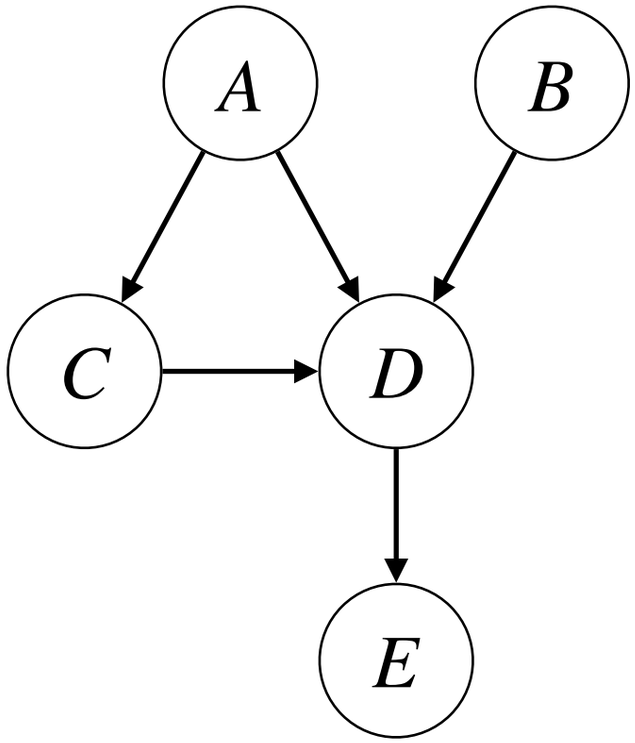
Review: Bayesian network

Review: Bayesian network



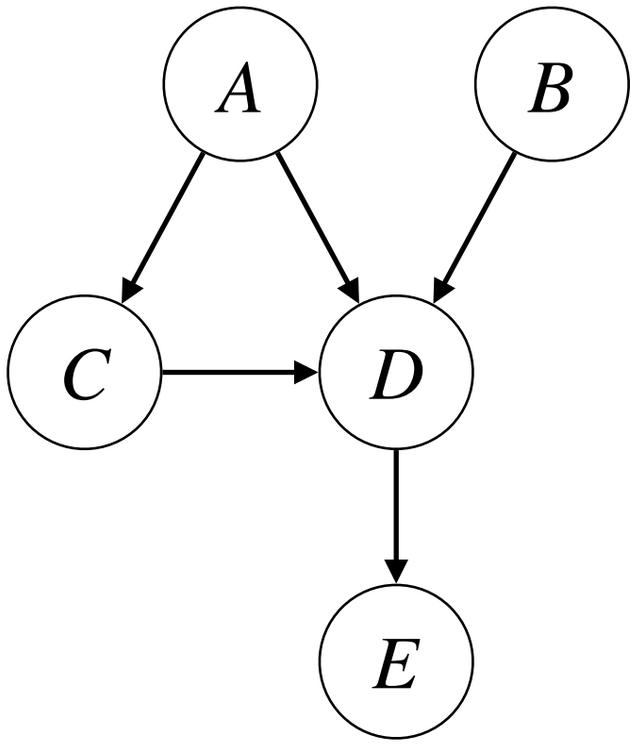
Review: Bayesian network

Directed acyclic graph (DAG)



Review: Bayesian network

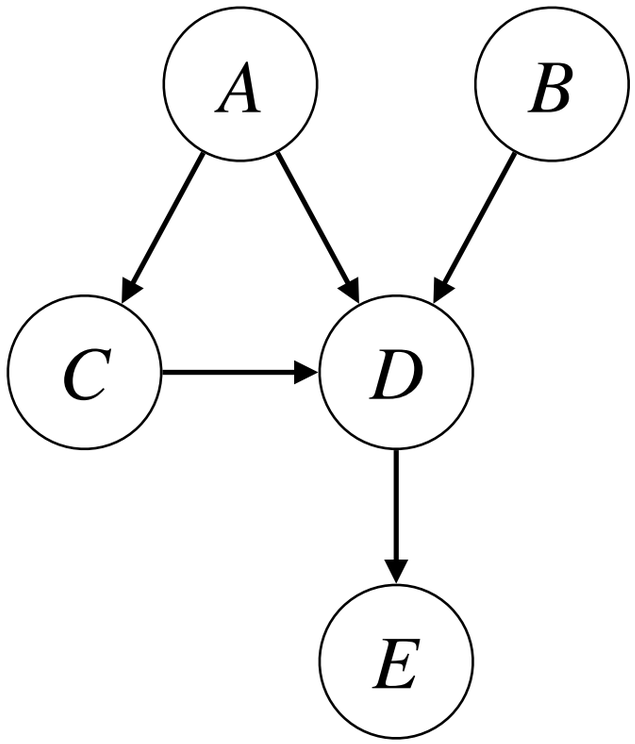
Directed acyclic graph (DAG)



Each node is a **random variable**.

Review: Bayesian network

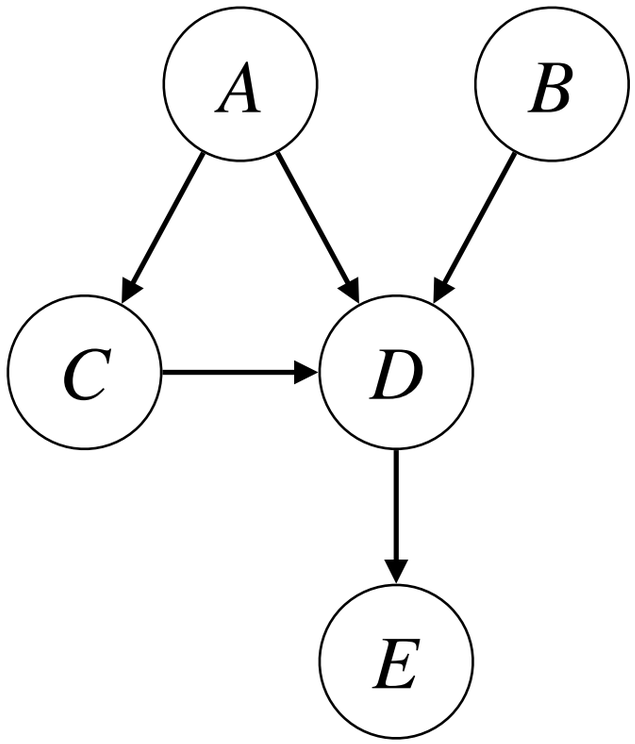
Directed acyclic graph (DAG)



Each node is a **random variable**.

Review: Bayesian network

Directed acyclic graph (DAG)

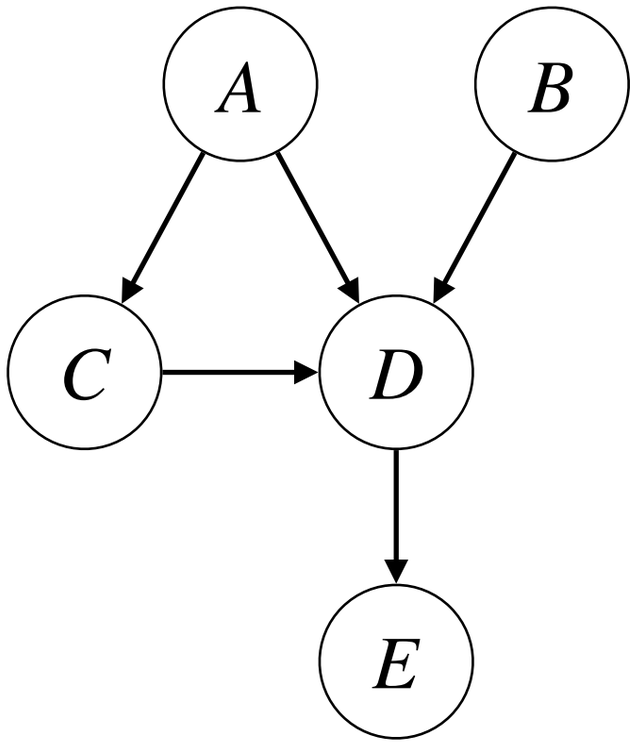


Each node is a **random variable**.

The graph defines a **joint distribution** over the variables.

Review: Bayesian network

Directed acyclic graph (DAG)

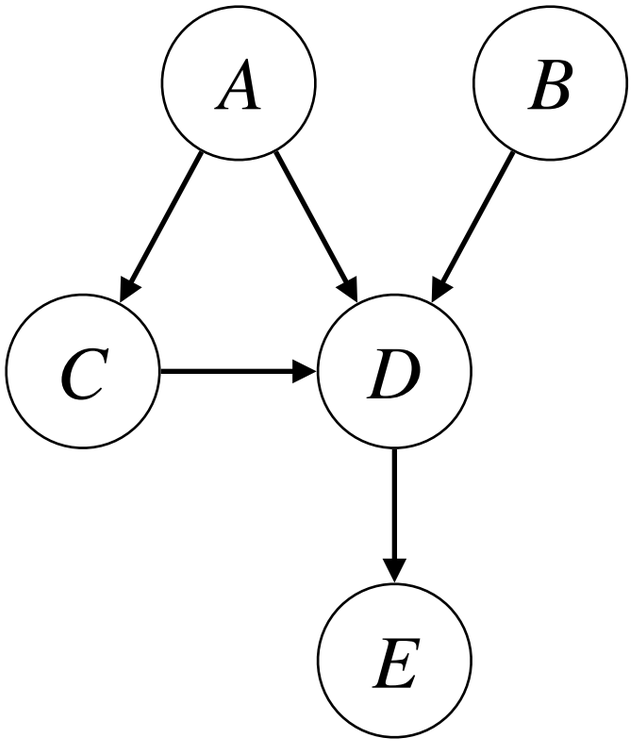


Each node is a **random variable**.

The graph defines a **joint distribution** over the variables.

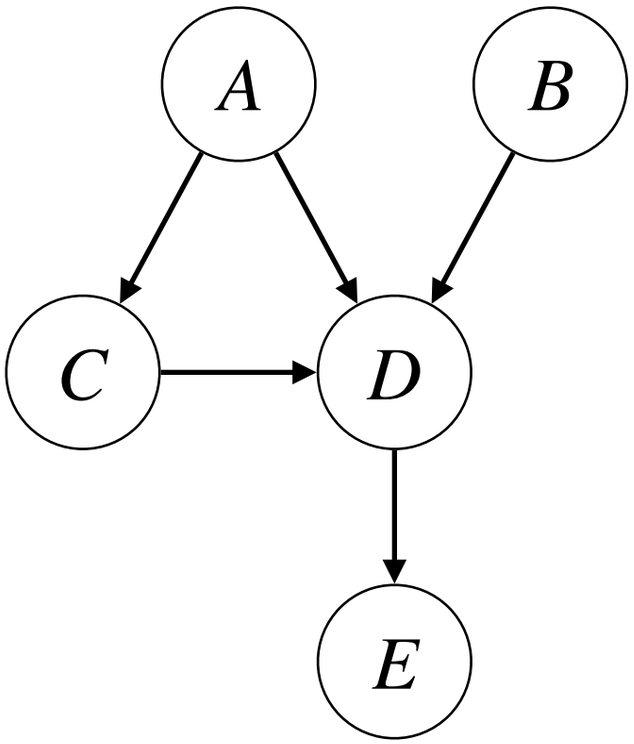
$$P(A, B, C, D, E)$$

Key property:
joint distribution = product of local distributions



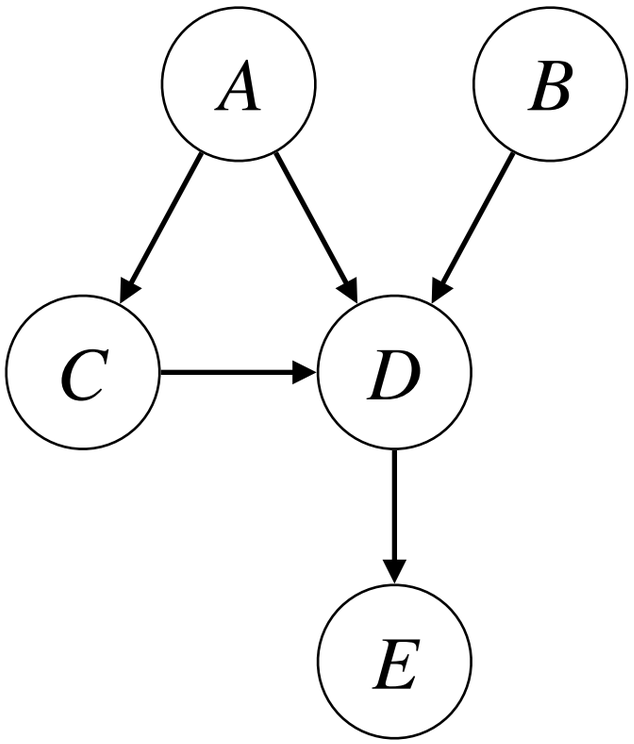
Key property:
joint distribution = product of local distributions

$$P(A, B, C, D, E) =$$



Key property:
joint distribution = product of local distributions

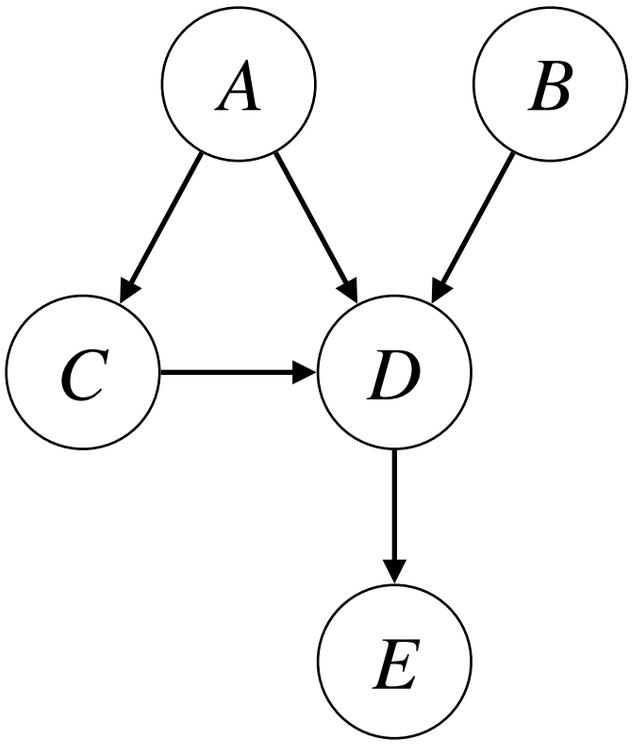
$$P(A, B, C, D, E) =$$



*product of **5** local distributions*

Key property:
joint distribution = product of local distributions

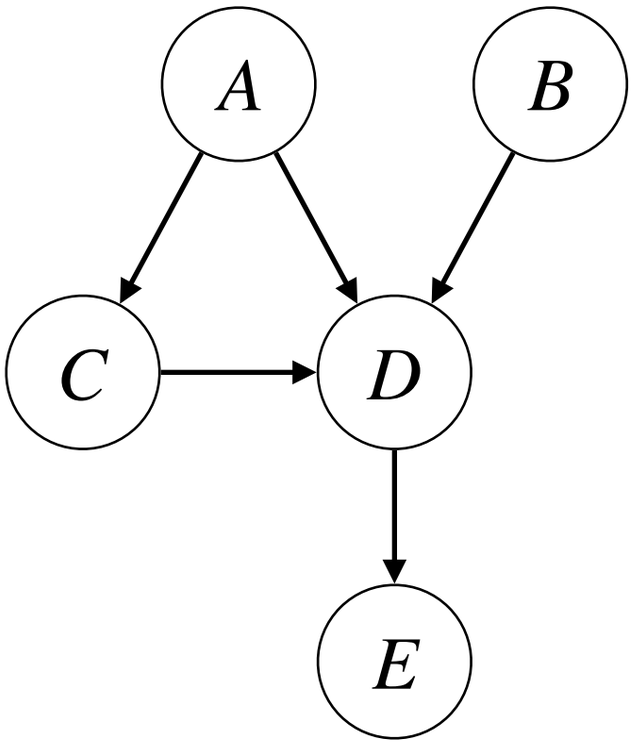
$$P(A, B, C, D, E) =$$



*product of **5** local distributions*

Key property:
joint distribution = product of local distributions

$$P(A, B, C, D, E) =$$

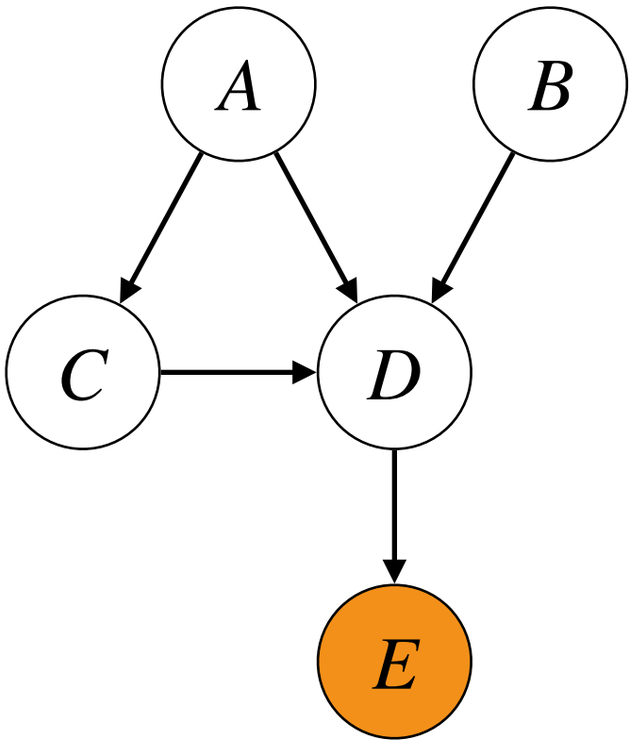


*product of **5** local distributions*

(one for each node in the graph)

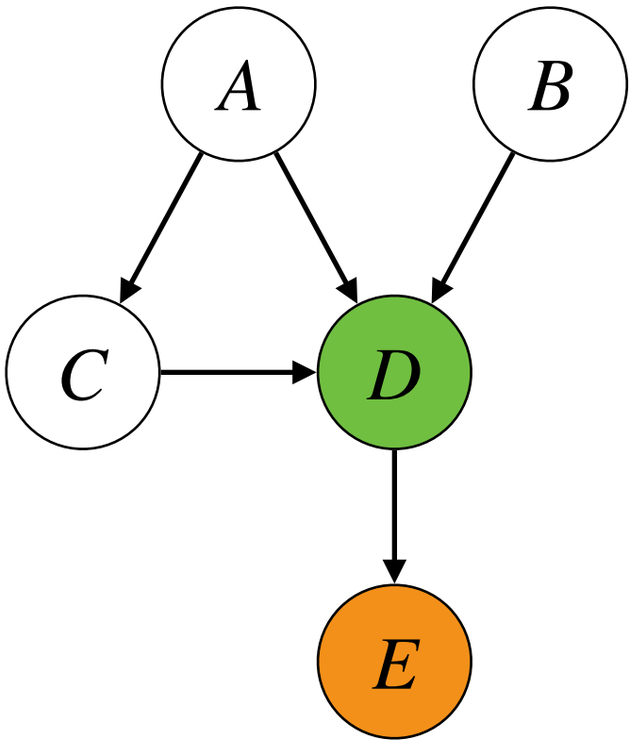
Key property:
joint distribution = product of local distributions

$$P(A, B, C, D, E) =$$



Key property:
joint distribution = product of local distributions

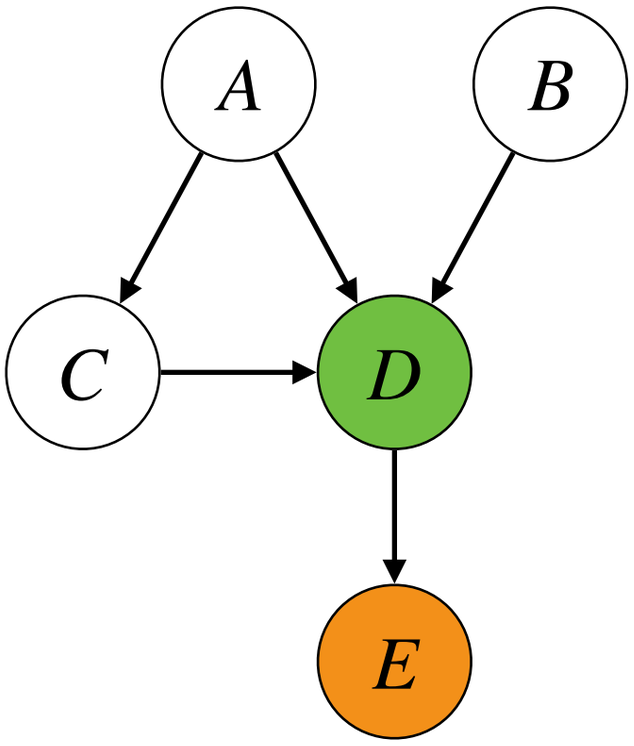
$$P(A, B, C, D, E) =$$



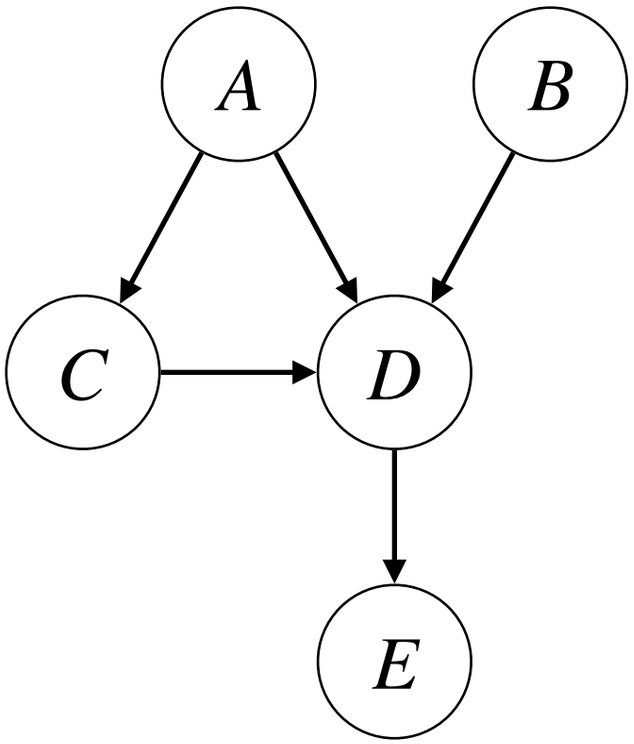
Key property:
joint distribution = product of local distributions

$$P(A, B, C, D, E) =$$

$$P(E \mid D)$$



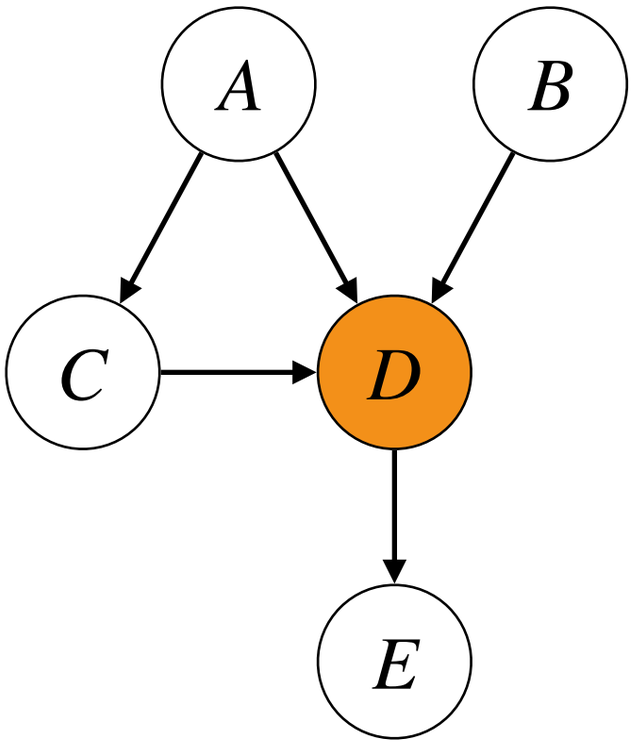
Key property:
joint distribution = product of local distributions



$$P(A, B, C, D, E) =$$

$$P(\mathbf{E} \mid \mathbf{D})$$

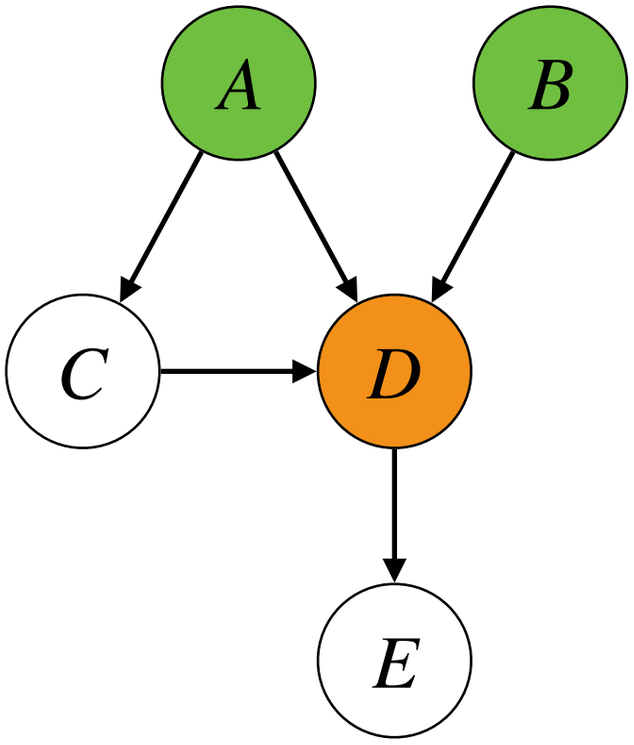
Key property:
joint distribution = product of local distributions



$$P(A, B, C, D, E) =$$

$$P(\boldsymbol{E} \mid \boldsymbol{D})$$

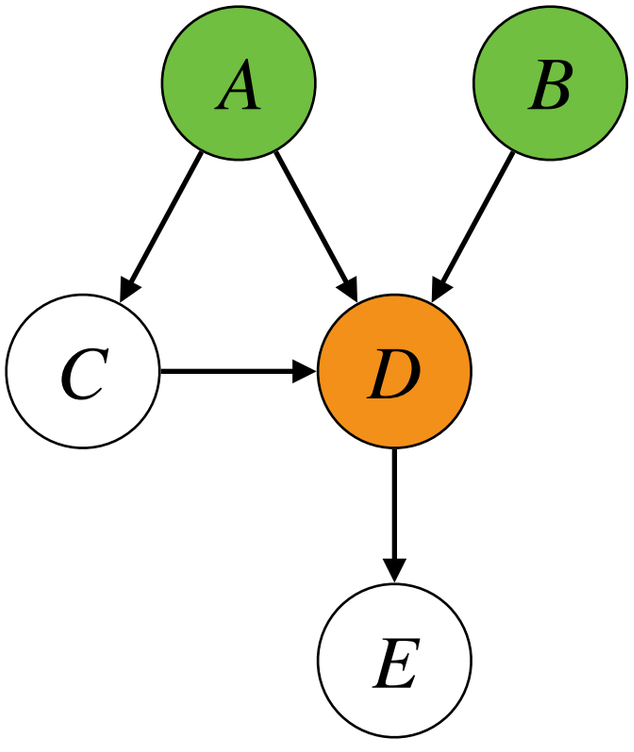
Key property:
joint distribution = product of local distributions



$$P(A, B, C, D, E) =$$

$$P(\boldsymbol{E} \mid \boldsymbol{D})$$

Key property:
joint distribution = product of local distributions

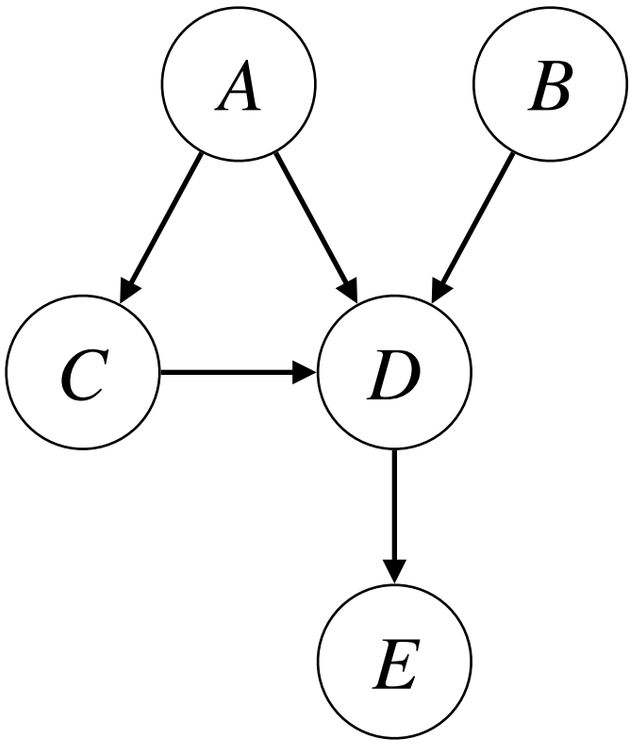


$$P(A, B, C, D, E) =$$

$$P(\boldsymbol{E} \mid \boldsymbol{D})$$

$$P(\boldsymbol{D} \mid \boldsymbol{A}, \boldsymbol{B})$$

Key property:
joint distribution = product of local distributions

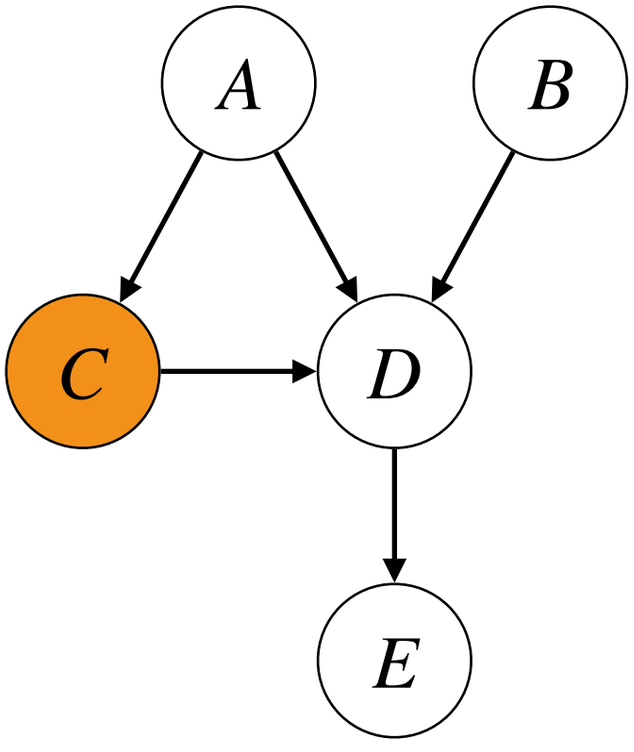


$$P(A, B, C, D, E) =$$

$$P(\boldsymbol{E} \mid \boldsymbol{D})$$

$$P(\boldsymbol{D} \mid \boldsymbol{A}, \boldsymbol{B})$$

Key property:
joint distribution = product of local distributions

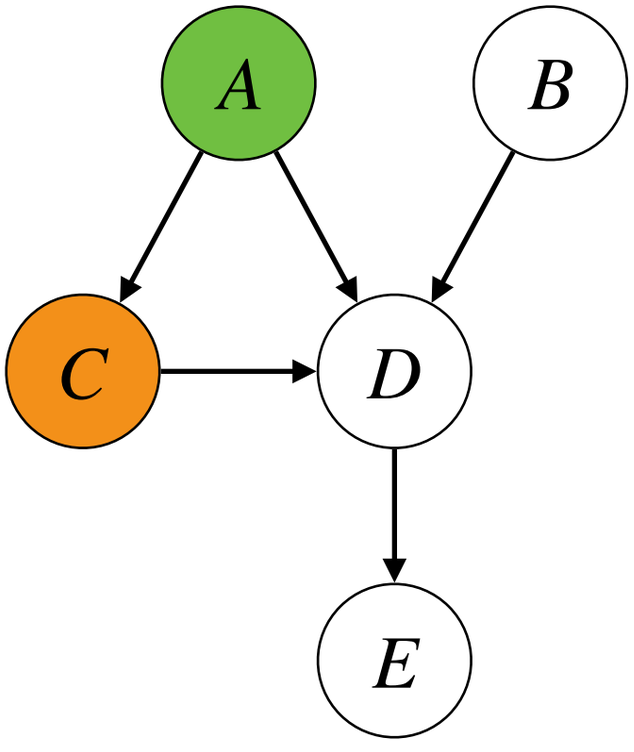


$$P(A, B, C, D, E) =$$

$$P(\boldsymbol{E} \mid \boldsymbol{D})$$

$$P(\boldsymbol{D} \mid \boldsymbol{A}, \boldsymbol{B})$$

Key property:
joint distribution = product of local distributions

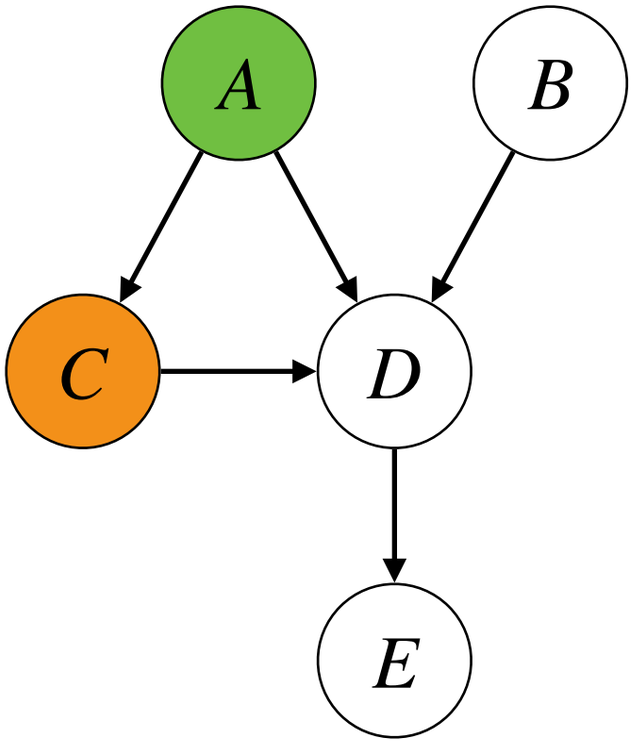


$$P(A, B, C, D, E) =$$

$$P(\boldsymbol{E} \mid \boldsymbol{D})$$

$$P(\boldsymbol{D} \mid \boldsymbol{A}, \boldsymbol{B})$$

Key property:
joint distribution = product of local distributions



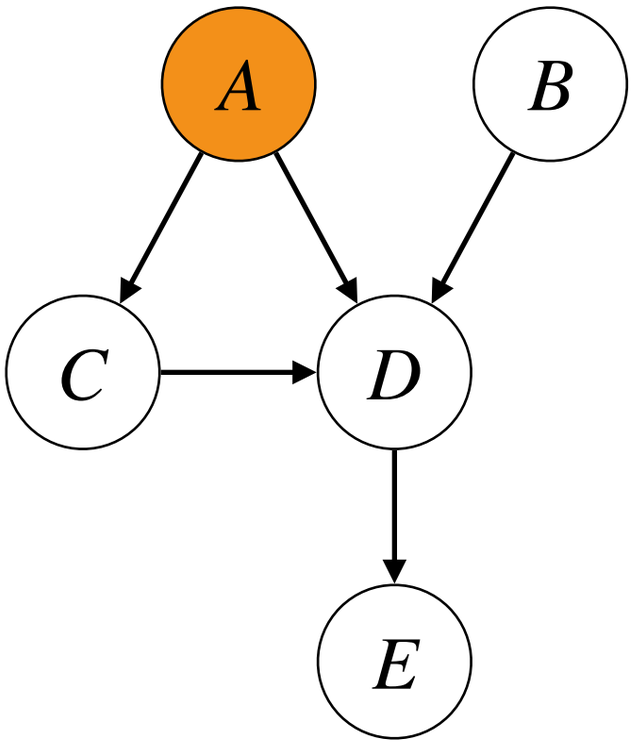
$$P(A, B, C, D, E) =$$

$$P(\boldsymbol{E} \mid \boldsymbol{D})$$

$$P(\boldsymbol{D} \mid \boldsymbol{A}, \boldsymbol{B})$$

$$P(\boldsymbol{C} \mid \boldsymbol{A})$$

Key property:
joint distribution = product of local distributions



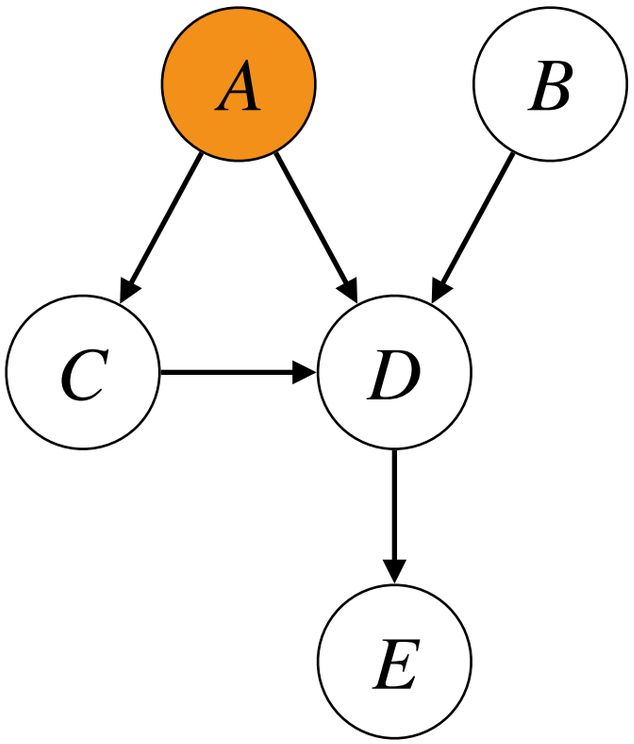
$$P(A, B, C, D, E) =$$

$$P(E \mid D)$$

$$P(D \mid A, B)$$

$$P(C \mid A)$$

Key property:
joint distribution = product of local distributions



$$P(A, B, C, D, E) =$$

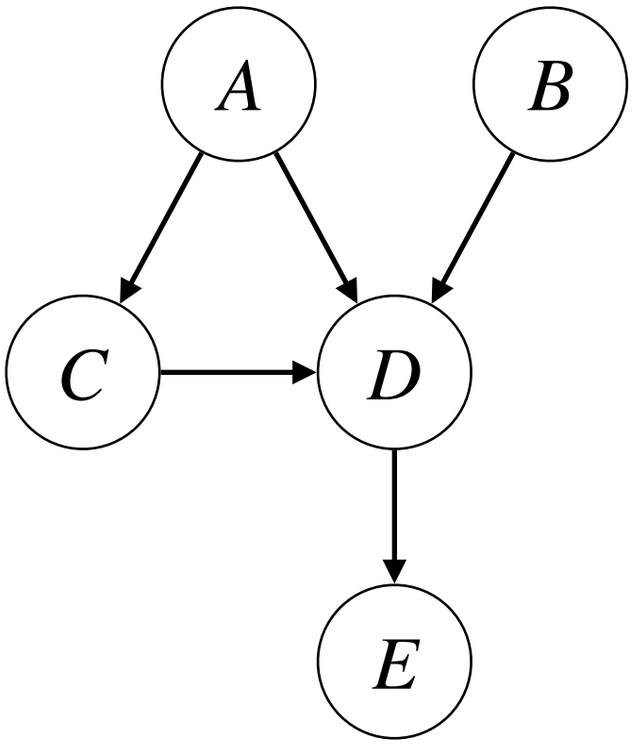
$$P(E \mid D)$$

$$P(D \mid A, B)$$

$$P(C \mid A)$$

$$P(A)$$

Key property:
joint distribution = product of local distributions



$$P(A, B, C, D, E) =$$

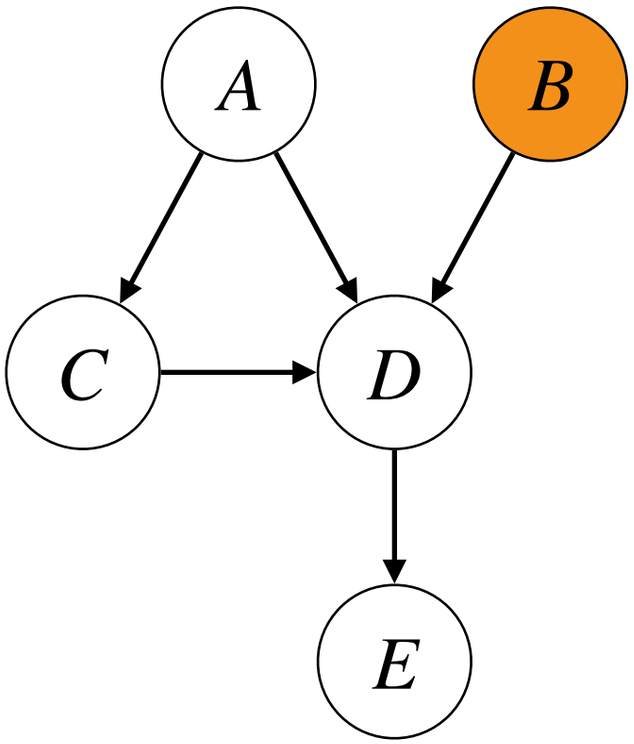
$$P(\boldsymbol{E} \mid \boldsymbol{D})$$

$$P(\boldsymbol{D} \mid \boldsymbol{A}, \boldsymbol{B})$$

$$P(\boldsymbol{C} \mid \boldsymbol{A})$$

$$P(\boldsymbol{A})$$

Key property:
joint distribution = product of local distributions



$$P(A, B, C, D, E) =$$

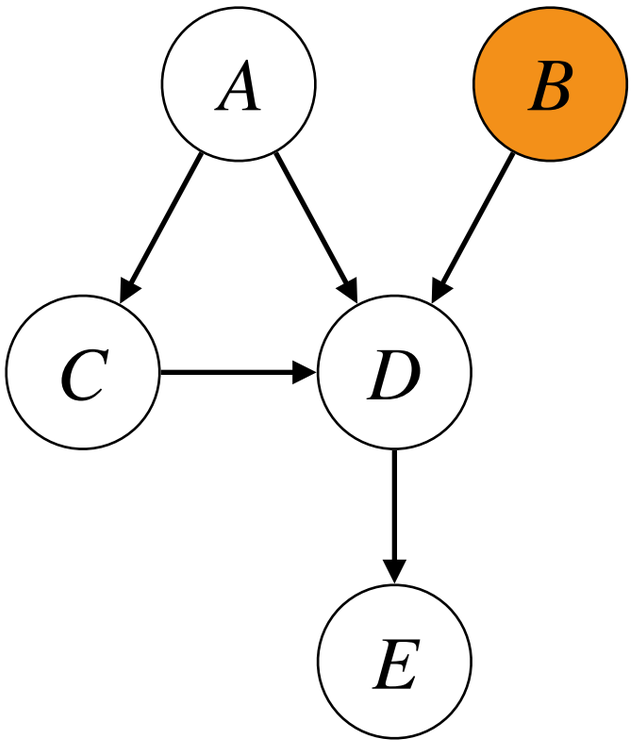
$$P(E \mid D)$$

$$P(D \mid A, B)$$

$$P(C \mid A)$$

$$P(A)$$

Key property:
joint distribution = product of local distributions



$$P(A, B, C, D, E) =$$

$$P(E \mid D)$$

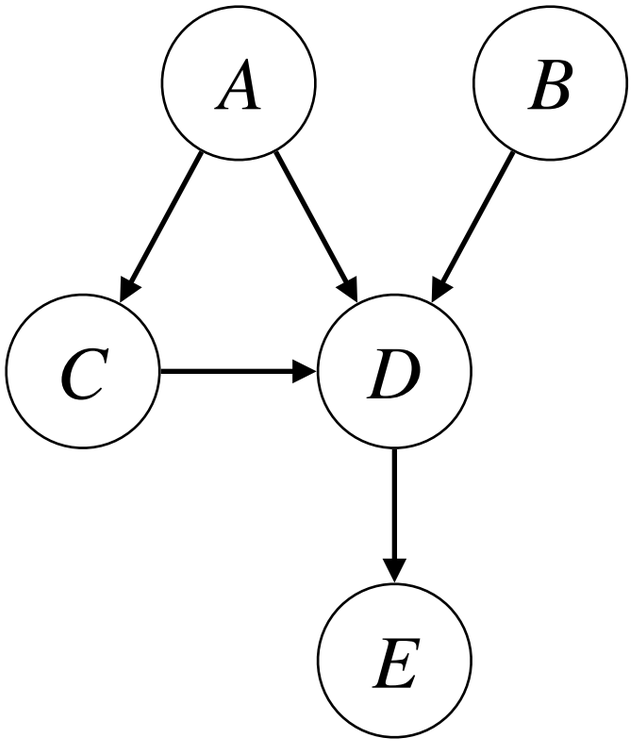
$$P(D \mid A, B)$$

$$P(C \mid A)$$

$$P(A)$$

$$P(B)$$

Key property:
joint distribution = product of local distributions



$$P(A, B, C, D, E) =$$

$$P(E \mid D)$$
$$P(D \mid A, B)$$
$$P(C \mid A)$$
$$P(A)$$
$$P(B)$$

product

Learning the parameters of a Bayesian network

data = {(d, 4), (d, 4), (d, 5), (c, 1), (c, 5)}

model = $p_G(g)\ \ p_R(r \mid g)$

data = {(d, 4), (d, 4), (d, 5), (c, 1), (c, 5)}

model = $p_G(g)$ $p_R(r \mid g)$

data = {(d, 4), (d, 4), (d, 5), (c, 1), (c, 5)}

model = $p_G(g)$ $p_R(r \mid g)$

g	$p_G(g)$
d	
c	

data = {(**d**, 4), (**d**, 4), (**d**, 5), (**c**, 1), (**c**, 5)}

model = $p_G(g)$ $p_R(r \mid g)$

g	$p_G(g)$
d	
c	

data = {(**d**, 4), (**d**, 4), (**d**, 5), (**c**, 1), (**c**, 5)}

model = $p_G(g)$ $p_R(r \mid g)$

g	$p_G(g)$
d	**3 / 5**
c	

data = {(**d**, 4), (**d**, 4), (**d**, 5), (**c**, 1), (**c**, 5)}

model = $p_G(g)$ $p_R(r \mid g)$

g	$p_G(g)$
d	**3 / 5**
c	**2 / 5**

data = {(d, 4), (d, 4), (d, 5), (c, 1), (c, 5)}

model = $p_G(g)$ $p_R(r \mid g)$

data = {(d, 4), (d, 4), (d, 5), (c, 1), (c, 5)}

model = $p_G(g)$ $p_R(r \mid g)$

data = {(d, 4), (d, 4), (d, 5), (c, 1), (c, 5)}

model = $p_G(g)$ $p_R(r \mid g)$

g	r	$p_R(r \mid g)$
d	1	
d	2	
d	3	
d	4	
d	5	
c	1	
c	2	
c	3	
c	4	
c	5	

data = {(d, 4), (d, 4), (d, 5), (c, 1), (c, 5)}

model = $p_G(g)$ $p_R(r \mid g)$

g	r	$p_R(r \mid g)$
d	1	
d	2	
d	3	
d	4	
d	5	
c	1	
c	2	
c	3	
c	4	
c	5	

data = {**(d, 4), (d, 4), (d, 5)**, (c, 1), (c, 5)}

model = $p_G(g)$ $p_R(r \mid g)$

g	r	$p_R(r \mid g)$
d	1	
d	2	
d	3	
d	4	
d	5	
c	1	
c	2	
c	3	
c	4	
c	5	

data = {**(d, 4), (d, 4), (d, 5)**, (c, 1), (c, 5)}

model = $p_G(g)$ $p_R(r \mid g)$

g	r	$p_R(r \mid g)$
d	1	
d	2	
d	3	
d	4	**2 / 3**
d	5	**1 / 3**
c	1	
c	2	
c	3	
c	4	
c	5	

data = {**(d, 4), (d, 4), (d, 5)**, (c, 1), (c, 5)}

model = $p_G(g)$ $p_R(r \mid g)$

g	r	$p_R(r \mid g)$
d	1	**0**
d	2	**0**
d	3	**0**
d	4	**2 / 3**
d	5	**1 / 3**
c	1	
c	2	
c	3	
c	4	
c	5	

data = {(d, 4), (d, 4), (d, 5), (c, 1), (c, 5)}

model = $p_G(g)$ $p_R(r \mid g)$

g	r	$p_R(r \mid g)$
d	1	0
d	2	0
d	3	0
d	4	2 / 3
d	5	1 / 3
c	1	
c	2	
c	3	
c	4	
c	5	

data = {(d, 4), (d, 4), (d, 5), (c, 1), (c, 5)}

model = $p_G(g)$ $p_R(r \mid g)$

g	r	$p_R(r \mid g)$
d	1	0
d	2	0
d	3	0
d	4	2 / 3
d	5	1 / 3
c	1	
c	2	
c	3	
c	4	
c	5	

data = {(d, 4), (d, 4), (d, 5), **(c, 1), (c, 5)**}

model = $p_G(g)$ $p_R(r \mid g)$

g	r	$p_R(r \mid g)$
d	1	0
d	2	0
d	3	0
d	4	2 / 3
d	5	1 / 3
c	1	
c	2	
c	3	
c	4	
c	5	

data = {(d, 4), (d, 4), (d, 5), **(c, 1), (c, 5)**}

model = $p_G(g)$ $p_R(r \mid g)$

g	r	$p_R(r \mid g)$
d	1	0
d	2	0
d	3	0
d	4	2 / 3
d	5	1 / 3
c	1	**1 / 2**
c	2	**0**
c	3	**0**
c	4	**0**
c	5	**1 / 2**

data =

(d, 0, 3)
(d, 1, 5)
(c, 0, 1)
(c, 0, 5)
(c, 1, 4)

$p_R(r \mid g, a)$

g	a	r	$p_R(r \mid g, a)$
d	0	1	
d	0	2	
d	0	3	
d	0	4	
d	0	5	
d	1	1	
d	1	2	
d	1	3	
d	1	4	
d	1	5	
c	0	1	
c	0	2	
c	0	3	
c	0	4	
c	0	5	
c	1	1	
c	1	2	
c	1	3	
c	1	4	
c	1	5	

data =

(d, 0, 3)
(d, 1, 5)
(c, 0, 1)
(c, 0, 5)
(c, 1, 4)

$p_R(r \mid g, a)$

g	a	r	$p_R(r \mid g, a)$
d	0	1	
d	0	2	
d	0	3	
d	0	4	
d	0	5	
d	1	1	
d	1	2	
d	1	3	
d	1	4	
d	1	5	
c	0	1	
c	0	2	
c	0	3	
c	0	4	
c	0	5	
c	1	1	
c	1	2	
c	1	3	
c	1	4	
c	1	5	

data =

(d, 0, 3)
(d, 1, 5)
(c, 0, 1)
(c, 0, 5)
(c, 0, 4)

$p_R(r \mid g, a)$

g	*a*	*r*	$p_R(r \mid g, a)$
d	0	1	
		2	
		3	
		4	
		5	
d	1	1	
d	1	2	
d	1	3	
d	1	4	
d	1	5	
c	0	1	
c	0	2	
c	0	3	
c	0	4	
c	0	5	
c	1	1	
c	1	2	
c	1	3	
c	1	4	
c	1	5	

data =

(d, 0, 3)
(d, 1, 5)
(c, 0, 1)
(c, 0, 5)
(c, 1, 4)

$p_R(r \mid g, a)$

g	a	r	$p_R(r \mid g, a)$
d	0	1	
		2	
		3	
		4	
		5	
d	1	1	
		2	
		3	
		4	
		5	
c	0	1	
c	0	2	
c	0	3	
c	0	4	
c	0	5	
c	1	1	
c	1	2	
c	1	3	
c	1	4	
c	1	5	

data =

(d, 0, 3)
(d, 1, 5)
(c, 0, 1)
(c, 0, 5)
(c, 0, 5)
(c, 1, 4)

$p_R(r \mid g, a)$

g	a	r	$p_R(r \mid g, a)$
d	0	1	
		2	
		3	
		4	
		5	
d	1	1	
		2	
		3	
		4	
		5	
c	0	1	
		2	
		3	
		4	
		5	
c	1	1	
c	1	2	
c	1	3	
c	1	4	
c	1	5	

data =

(d, 0, 3)
(d, 1, 5)
(c, 0, 1)
(c, 0, 5)
(c, 1, 4)

$p_R(r \mid g, a)$

g	a	r	$p_R(r \mid g, a)$
d	0	1	
		2	
		3	
		4	
		5	
d	1	1	
		2	
		3	
		4	
		5	
c	0	1	
		2	
		3	
		4	
		5	
c	1	1	
		2	
		3	
		4	
		5	

data =

(d, 0, 3)
(d, 1, 5)
(c, 0, 1)
(c, 0, 5)
(c, 1, 4)

$p_R(r \mid g, a)$

g	a	r	$p_R(r \mid g, a)$
d	0	1	
		2	
		3	
		4	
		5	
d	1	1	
		2	
		3	
		4	
		5	
c	0	1	
		2	
		3	
		4	
		5	
c	1	1	
		2	
		3	
		4	
		5	

data =

(d, 0, 3)
(d, 1, 5)
(c, 0, 1)
(c, 0, 5)
(c, 1, 4)

$p_R(r \mid g, a)$

g	a	r	$p_R(r \mid g, a)$
d	0	1	
		2	
		3	
		4	
		5	
d	1	1	
		2	
		3	
		4	
		5	
c	0	1	
		2	
		3	
		4	
		5	
c	1	1	
		2	
		3	
		4	
		5	

data =

(d, 0, 3)
(d, 1, 5)
(c, 0, 1)
(c, 0, 5)
(c, 1, 4)

$p_R(r \mid g, a)$

g	*a*	*r*	$p_R(r \mid g, a)$
d	0	1	**0**
		2	**0**
		3	**1**
		4	**0**
		5	**0**
d	1	1	
		2	
		3	
		4	
		5	
c	0	1	
		2	
		3	
		4	
		5	
c	1	1	
		2	
		3	
		4	
		5	

data =

(d, 0, 3)
(d, 1, 5)
(c, 0, 1)
(c, 0, 5)
(c, 1, 4)

$p_R(r \mid g, a)$

g	a	r	$p_R(r \mid g, a)$
d	0	1	0
		2	0
		3	1
		4	0
		5	0
d	1	1	
		2	
		3	
		4	
		5	
c	0	1	
		2	
		3	
		4	
		5	
c	1	1	
		2	
		3	
		4	
		5	

data =

(d, 0, 3)
(d, 1, 5)
(c, 0, 1)
(c, 0, 5)
(c, 1, 4)

$p_R(r \mid g, a)$

g	a	r	$p_R(r \mid g, a)$
d	0	1	0
		2	0
		3	1
		4	0
		5	0
d	1	1	
		2	
		3	
		4	
		5	
c	0	1	
		2	
		3	
		4	
		5	
c	1	1	
		2	
		3	
		4	
		5	

data =

(d, 0, 3)

(d, 1, 5)

(c, 0, 1)

(c, 0, 5)

(c, 1, 4)

$p_R(r \mid g, a)$

g	a	r	$p_R(r \mid g, a)$
d	0	1	0
		2	0
		3	1
		4	0
		5	0
d	1	1	
		2	
		3	
		4	
		5	
c	0	1	
		2	
		3	
		4	
		5	
c	1	1	
		2	
		3	
		4	
		5	

data =

(d, 0, 3)

(d, 1, 5)

(c, 0, 1)

(c, 0, 5)

(c, 1, 4)

$p_R(r \mid g, a)$

g	a	r	$p_R(r \mid g, a)$
d	0	1	0
		2	0
		3	1
		4	0
		5	0
d	1	1	**0**
		2	**0**
		3	**0**
		4	**0**
		5	**1**
c	0	1	
		2	
		3	
		4	
		5	
c	1	1	
		2	
		3	
		4	
		5	

data =

(d, 0, 3)
(d, 1, 5)
(c, 0, 1)
(c, 0, 5)
(c, 0, 5)
(c, 1, 4)

$p_R(r \mid g, a)$

g	a	r	$p_R(r \mid g, a)$
d	0	1	0
		2	0
		3	1
		4	0
		5	0
d	1	1	0
		2	0
		3	0
		4	0
		5	1
c	0	1	
		2	
		3	
		4	
		5	
c	1	1	
		2	
		3	
		4	
		5	

data =

(d, 0, 3)

(d, 1, 5)

(c, 0, 1)

(c, 0, 5)

(c, 1, 4)

$p_R(r \mid g, a)$

g	a	r	$p_R(r \mid g, a)$
d	0	1	0
		2	0
		3	1
		4	0
		5	0
d	1	1	0
		2	0
		3	0
		4	0
		5	1
c	0	1	
		2	
		3	
		4	
		5	
c	1	1	
		2	
		3	
		4	
		5	

data =

(d, 0, 3)

(d, 1, 5)

(c, 0, 1)

(c, 0, 5)

(c, 1, 4)

$p_R(r \mid g, a)$

g	a	r	$p_R(r \mid g, a)$
d	0	1	0
		2	0
		3	1
		4	0
		5	0
d	1	1	0
		2	0
		3	0
		4	0
		5	1
c	0	1	**1 / 2**
		2	**0**
		3	**0**
		4	**0**
		5	**1 / 2**
c	1	1	
		2	
		3	
		4	
		5	

data =

(d, 0, 3)
(d, 1, 5)
(c, 0, 1)
(c, 0, 5)
(c, 1, 4)

$p_R(r \mid g, a)$

g	a	r	$p_R(r \mid g, a)$
d	0	1	0
		2	0
		3	1
		4	0
		5	0
d	1	1	0
		2	0
		3	0
		4	0
		5	1
c	0	1	1 / 2
		2	0
		3	0
		4	0
		5	1 / 2
c	1	1	**0**
		2	**0**
		3	**0**
		4	**1**
		5	**0**

data =

(d, 0, 3)
(d, 1, 5)
(c, 0, 1)
(c, 0, 5)
(c, 1, 4)

$p_R(r \mid g, a)$

g	a	r	$p_R(r \mid g, a)$
d	0	1	0
		2	0
		3	1
		4	0
		5	0
d	1	1	0
		2	0
		3	0
		4	0
		5	1
c	0	1	1 / 2
		2	0
		3	0
		4	0
		5	1 / 2
c	1	1	0
		2	0
		3	0
		4	1
		5	0

data =

(d, 0, 3)
(d, 1, 5)
(c, 0, 1)
(c, 0, 5)
(c, 1, 4)

$p_R(r \mid g, a)$

input variables

g	*a*	*r*	$p_R(r \mid g, a)$
d	0	1	0
		2	0
		3	1
		4	0
		5	0
d	1	1	0
		2	0
		3	0
		4	0
		5	1
c	0	1	1 / 2
		2	0
		3	0
		4	0
		5	1 / 2
c	1	1	0
		2	0
		3	0
		4	1
		5	0

data =

(d, 0, 3)
(d, 1, 5)
(c, 0, 1)
(c, 0, 5)
(c, 1, 4)

$p_R(r \mid g, a)$

input variables

(**filter** the data)

g	*a*	*r*	$p_R(r \mid g, a)$
d	0	1	0
		2	0
		3	1
		4	0
		5	0
d	1	1	0
		2	0
		3	0
		4	0
		5	1
c	0	1	1 / 2
		2	0
		3	0
		4	0
		5	1 / 2
c	1	1	0
		2	0
		3	0
		4	1
		5	0

data =

(d, 0, 3)
(d, 1, 5)
(c, 0, 1)
(c, 0, 5)
(c, 1, 4)

$p_R(r \mid g, a)$

g	a	r	$p_R(r \mid g, a)$
d	0	1	0
		2	0
		3	1
		4	0
		5	0
d	1	1	0
		2	0
		3	0
		4	0
		5	1
c	0	1	1 / 2
		2	0
		3	0
		4	0
		5	1 / 2
c	1	1	0
		2	0
		3	0
		4	1
		5	0

data =

(d, 0, 3)
(d, 1, 5)
(c, 0, 1)
(c, 0, 5)
(c, 1, 4)

$p_R(r \mid g, a)$

g	a	r	$p_R(r \mid g, a)$
d	0	1	0
		2	0
		3	1
		4	0
		5	0
d	1	1	0
		2	0
		3	0
		4	0
		5	1
c	0	1	1 / 2
		2	0
		3	0
		4	0
		5	1 / 2
c	1	1	0
		2	0
		3	0
		4	1
		5	0

output variable

data =

(d, 0, 3)
(d, 1, 5)
(c, 0, 1)
(c, 0, 5)
(c, 1, 4)

$p_R(r \mid g, a)$

g	a	r	$p_R(r \mid g, a)$
d	0	1	0
		2	0
		3	1
		4	0
		5	0
d	1	1	0
		2	0
		3	0
		4	0
		5	1
c	0	1	1 / 2
		2	0
		3	0
		4	0
		5	1 / 2
c	1	1	0
		2	0
		3	0
		4	1
		5	0

data =

(d, 0, 3)
(d, 1, 5)
(c, 0, 1)
(c, 0, 5)
(c, 1, 4)

$p_R(r \mid g, a)$

each block sums to 1

g	*a*	*r*	$p_R(r \mid g, a)$
d	0	1	0
		2	0
		3	1
		4	0
		5	0
d	1	1	0
		2	0
		3	0
		4	0
		5	1
c	0	1	1 / 2
		2	0
		3	0
		4	0
		5	1 / 2
c	1	1	0
		2	0
		3	0
		4	1
		5	0

data =

(d, 0, 3)
(d, 1, 5)
(c, 0, 1)
(c, 0, 5)
(c, 1, 4)

$p_R(r \mid g, a)$

g	a	r	$p_R(r \mid g, a)$
d	0	1	0
		2	0
		3	1
		4	0
		5	0
d	1	1	0
		2	0
		3	0
		4	0
		5	1
c	0	1	1 / 2
		2	0
		3	0
		4	0
		5	1 / 2
c	1	1	0
		2	0
		3	0
		4	1
		5	0

data =

(d, 0, 3)
(d, 1, 5)
(c, 0, 1)
(c, 0, 5)
(c, 1, 4)

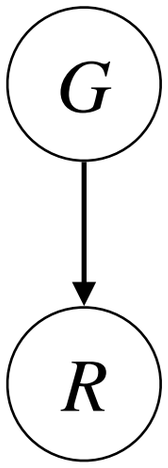
$p_R(r \mid g, a)$

g	a	r	$p_R(r \mid g, a)$
d	0	3	1
d	1	5	1
c	0	1	1 / 2
c	0	5	1 / 2
c	1	4	1

Handling vector-valued variables in a Bayesian network

What about vector-valued variables?

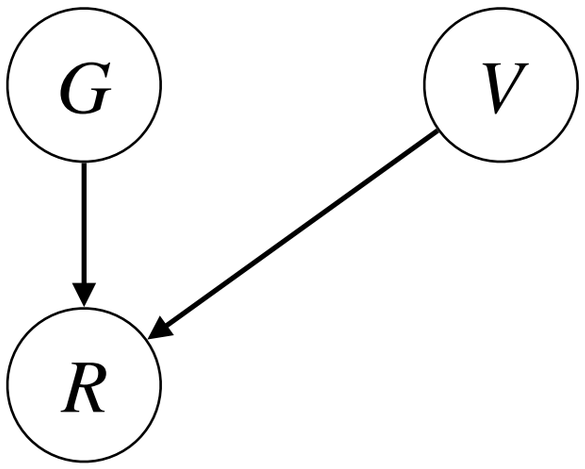
{drama, comedy}



{1, 2, 3, 4, 5}

What about vector-valued variables?

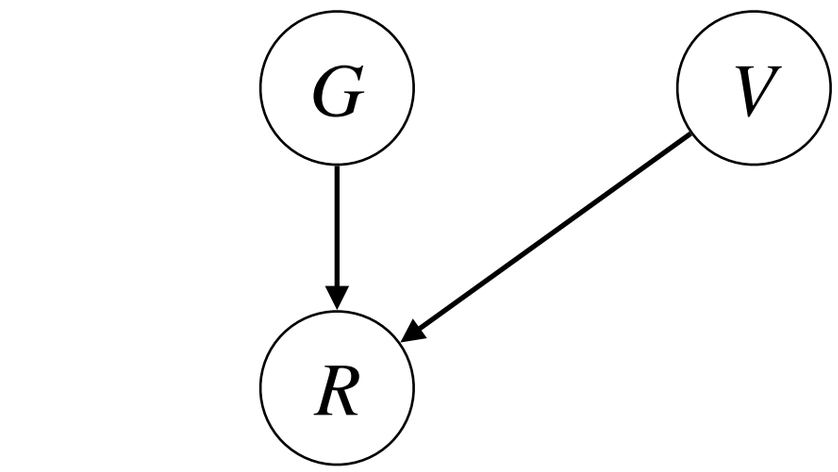
{drama, comedy}



{1, 2, 3, 4, 5}

What about vector-valued variables?

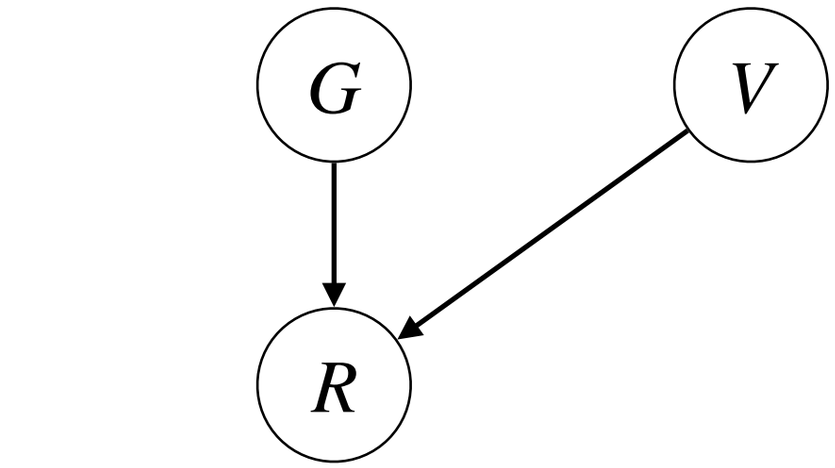
{drama, comedy}



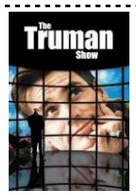
{1, 2, 3, 4, 5}

What about vector-valued variables?

{drama, comedy}

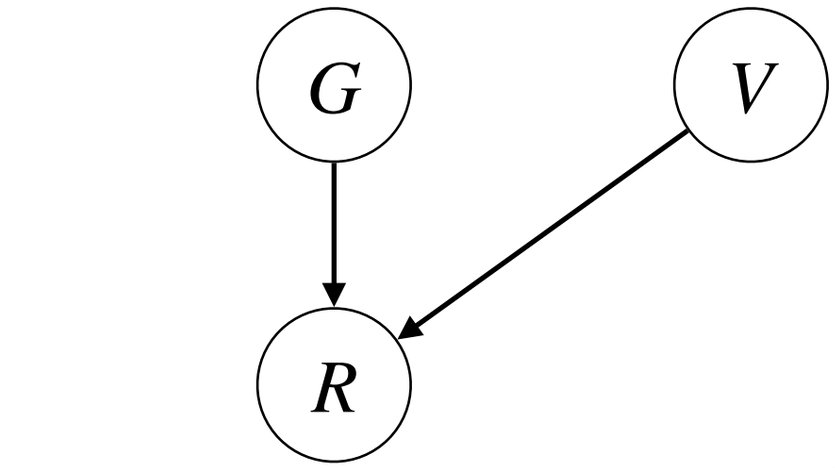


{1, 2, 3, 4, 5}

v	g	r	$p_R(r \mid v, g)$
	d	…	
	c	…	
	d	…	
	c	…	
…	…	…	…

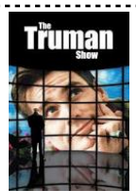
What about vector-valued variables?

{drama, comedy}



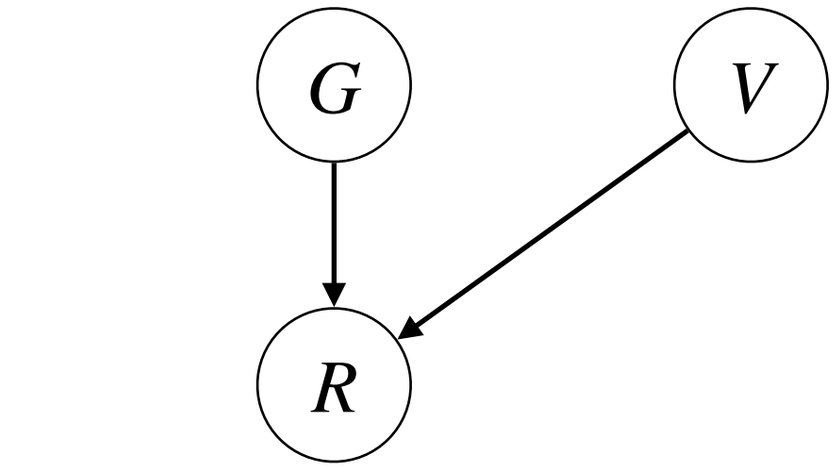
{1, 2, 3, 4, 5}

256^2 pixels

v	g	r	$p_R(r \mid v, g)$
	d	…	
	c	…	
	d	…	
	c	…	
…	…	…	…

What about vector-valued variables?

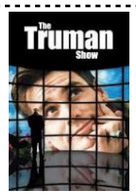
{drama, comedy}



{1, 2, 3, 4, 5}

256^2 pixels

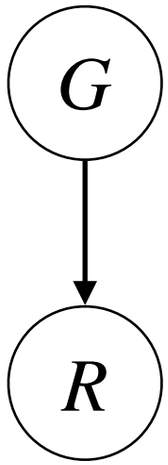
2^{65536} possible images

v	g	r	$p_R(r \mid v, g)$
	d	…	
	c	…	
	d	…	
	c	…	
…	…	…	…

What about vector-valued variables?

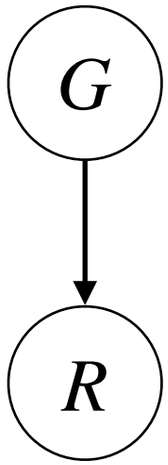
What about vector-valued variables?

Model



What about vector-valued variables?

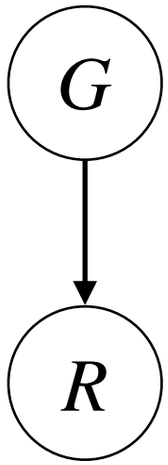
Model **Data**



{(d, 1), (c, 5), … }

What about vector-valued variables?

Model **Data**

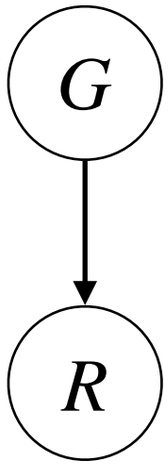


{(d, 1), (c, 5), … }

counting

What about vector-valued variables?

Model **Data** **Parameters**



{(d, 1), (c, 5), … }

counting

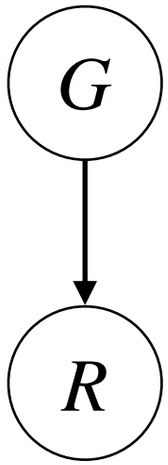
$\theta =$

g	$p_G(g)$
d	3/5
c	2/5

g	r	$p_R(r \mid g)$
d	4	2/3
d	5	1/3
c	1	1/2
c	5	1/2

What about vector-valued variables?

Model



Data

{(d, 1), (c, 5), … }

counting

Parameters

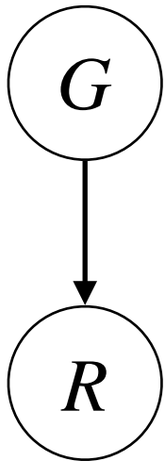
$\theta =$

g	$p_G(g)$
d	3/5
c	2/5

g	r	$p_R(r \mid g)$
d	4	2/3
d	5	1/3
c	1	1/2
c	5	1/2

What about vector-valued variables?

Model



Data

{(d, 1), (c, 5), … }

counting

Parameters

$\theta =$

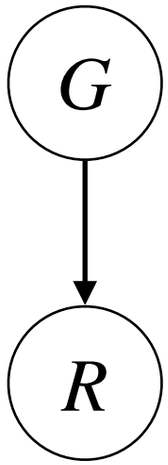
g	$p_G(g)$
d	3/5
c	2/5

g	r	$p_R(r \mid g)$
d	4	2/3
d	5	1/3
c	1	1/2
c	5	1/2

More general principle:

What about vector-valued variables?

Model



Data

{(d, 1), (c, 5), … }

counting

Parameters

$\theta =$

g	$p_G(g)$
d	3/5
c	2/5

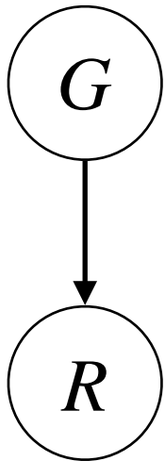
g	r	$p_R(r \mid g)$
d	4	2/3
d	5	1/3
c	1	1/2
c	5	1/2

More general principle:

$$\prod_{(g,r)\in\mathcal{D}_{\text{train}}} p_G(g)\, p_R(r \mid g)$$

What about vector-valued variables?

Model



Data

{(d, 1), (c, 5), … }

counting

Parameters

$\theta =$

g	$p_G(g)$
d	3/5
c	2/5

g	r	$p_R(r \mid g)$
d	4	2/3
d	5	1/3
c	1	1/2
c	5	1/2

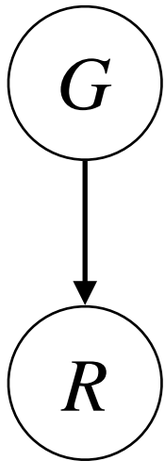
More general principle:

$$\prod_{(g,r)\in\mathcal{D}_{\text{train}}} p_G(g)\, p_R(r \mid g)$$

likelihood of the data

What about vector-valued variables?

Model



Data

{(d, 1), (c, 5), … }

counting

Parameters

$\theta =$

g	$p_G(g)$
d	3/5
c	2/5

g	r	$p_R(r \mid g)$
d	4	2/3
d	5	1/3
c	1	1/2
c	5	1/2

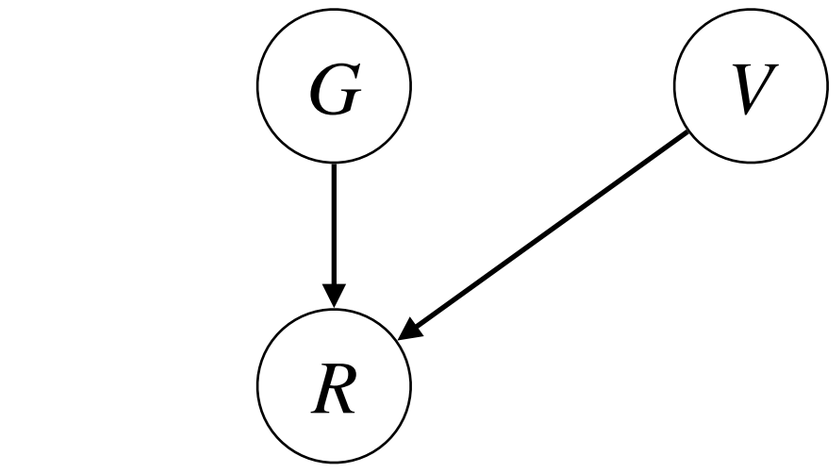
More general principle:

$$\theta = \underset{p_G,\, p_R}{\operatorname{argmax}} \prod_{(g,r)\in\mathcal{D}_{\text{train}}} p_G(g)\, p_R(r \mid g)$$

likelihood of the data

What about vector-valued variables?

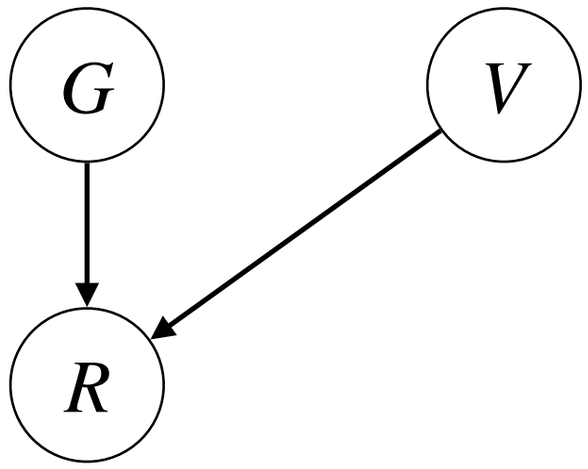
{drama, comedy}



{1, 2, 3, 4, 5}

What about vector-valued variables?

{drama, comedy}

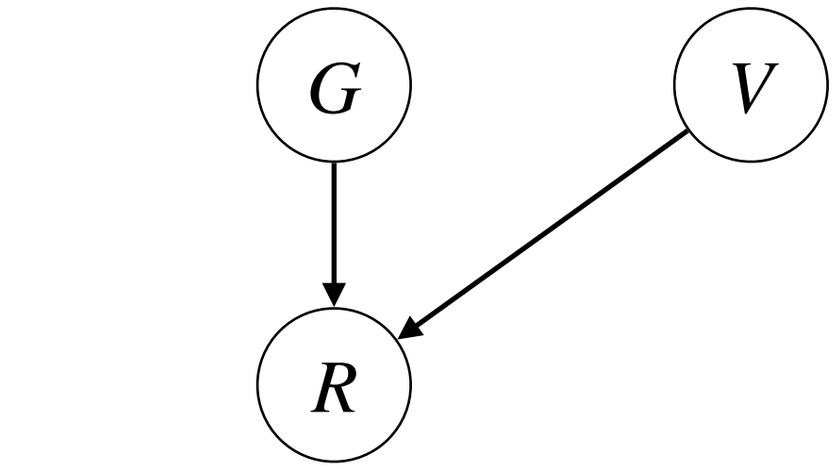


{1, 2, 3, 4, 5}

$$\prod_{(g,r,v)\in\mathcal{D}_{\text{train}}} p_G(g)\, p_V(v)\, p_R(r \mid g, v)$$

What about vector-valued variables?

{drama, comedy}



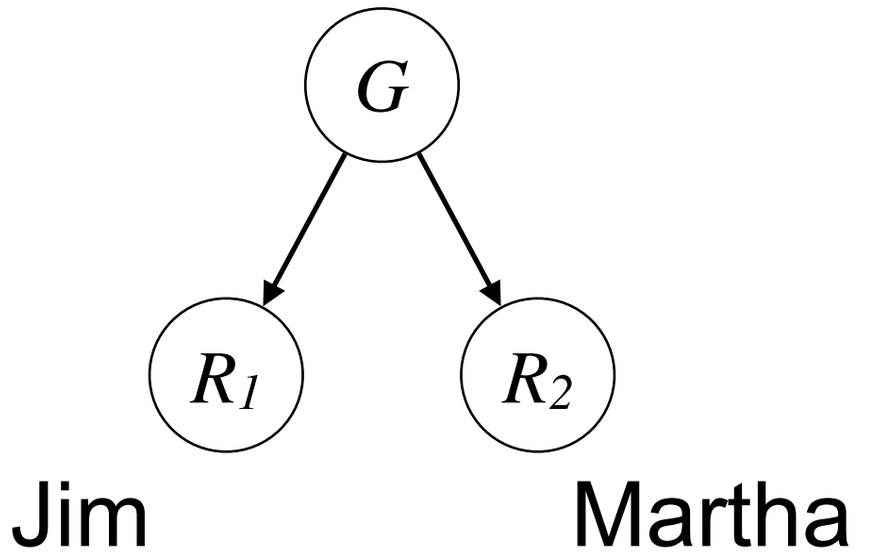
{1, 2, 3, 4, 5}

$$\prod_{(g,r,v)\in\mathcal{D}_{\text{train}}} p_G(g)\, p_V(v)\, p_R(r \mid g, v)$$

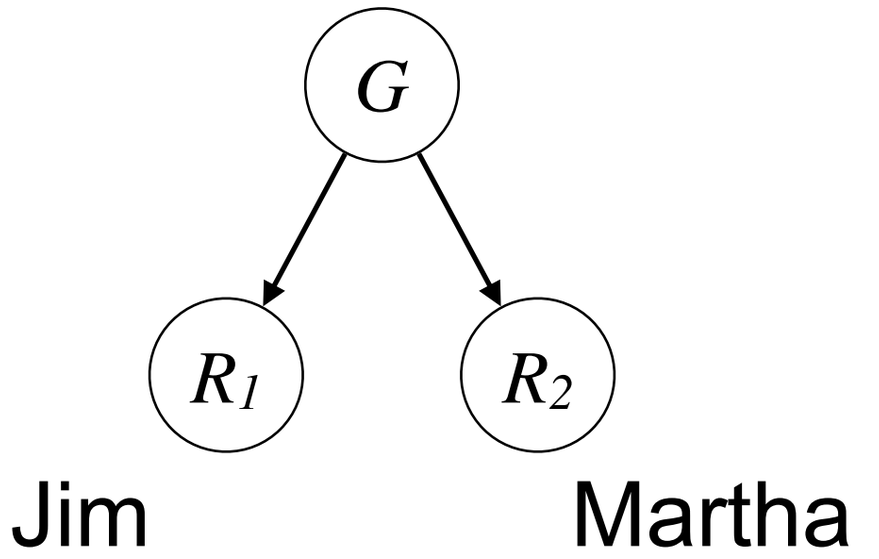
any probabilistic model
(e.g. neural network)

Parameter sharing in a Bayesian network

Parameter sharing

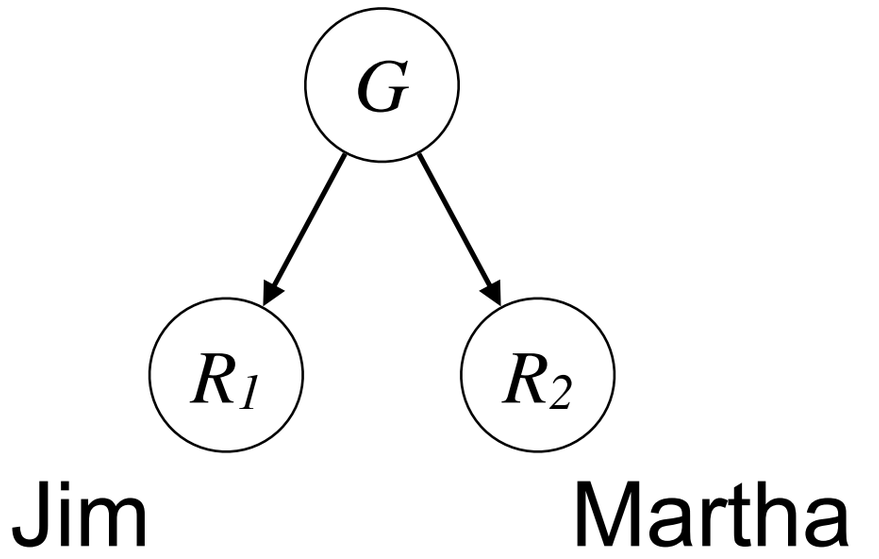


Parameter sharing



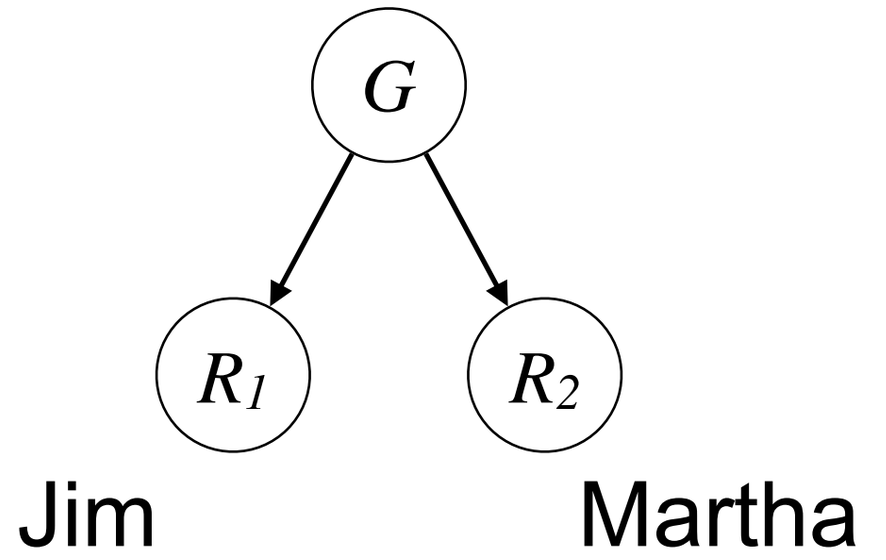
We think Jim and Martha have
the exact same preferences

Parameter sharing



We think Jim and Martha have
the exact same preferences

Parameter sharing



We think Jim and Martha have
the exact same preferences

Treat Jim's data and Martha's data
as if they came from one person

Parameter sharing

Parameter sharing

Jim's data

g	*r*	**count**
d	1	
	2	
	3	
	4	2
	5	1
c	1	1
	2	
	3	
	4	
	5	1

Parameter sharing

Jim's data

g	*r*	**count**
	1	
	2	
d	3	
	4	2
	5	1
	1	1
	2	
c	3	
	4	
	5	1

+

Martha's data

g	*r*	**count**
	1	
	2	
d	3	1
	4	1
	5	1
	1	
	2	1
c	3	
	4	1
	5	

Parameter sharing

Jim's data

g	*r*	**count**
	1	
	2	
d	3	
	4	2
	5	1
	1	1
	2	
c	3	
	4	
	5	1

+

Martha's data

g	*r*	**count**
	1	
	2	
d	3	1
	4	1
	5	1
	1	
	2	1
c	3	
	4	1
	5	

=

Merged data

g	*r*	**count**
	1	
	2	
d	3	**1**
	4	**3**
	5	**2**
	1	**1**
	2	**1**
c	3	
	4	**1**
	5	**1**

Parameter sharing

Merged data

g	*r*	**count**
d	1	
	2	
	3	**1**
	4	**3**
	5	**2**
c	1	**1**
	2	**1**
	3	
	4	**1**
	5	**1**

Parameter sharing

Merged data

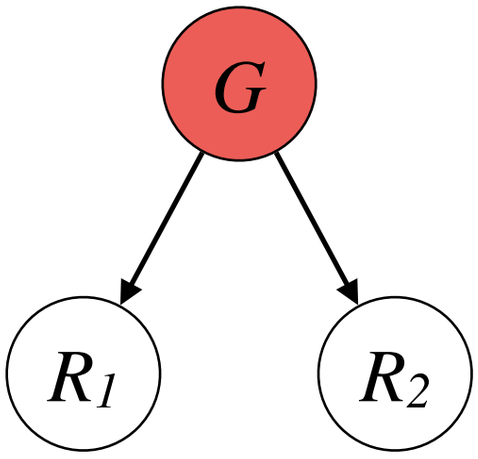
g	*r*	**count**
	1	
	2	
d	3	**1**
	4	**3**
	5	**2**
	1	**1**
	2	**1**
c	3	
	4	**1**
	5	**1**

Normalized

g	*r*	***p(r | g)***
	1	
	2	
d	3	**1 / 6**
	4	**3 / 6**
	5	**2 / 6**
	1	**1 / 4**
	2	**1 / 4**
c	3	
	4	**1 / 4**
	5	**1 / 4**

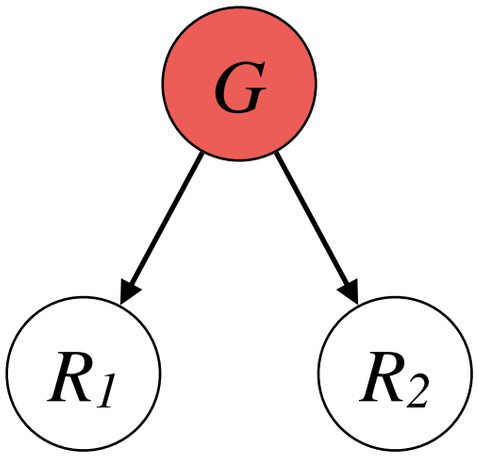
Expectation Maximization

Expectation maximization



$$\text{data} = \{(?, 4, 5), (?, 4, 4), (?, 5, 3), (?, 1, 2), (?, 5, 4)\}$$

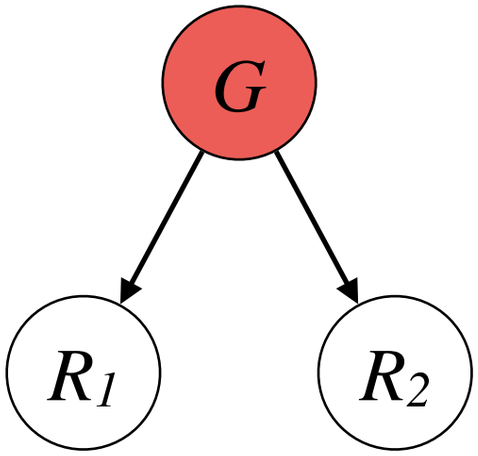
Expectation maximization



$$\text{data} = \{(?, 4, 5), (?, 4, 4), (?, 5, 3), (?, 1, 2), (?, 5, 4)\}$$

```
For each data point:
  "guess" the value of G
```

Expectation maximization

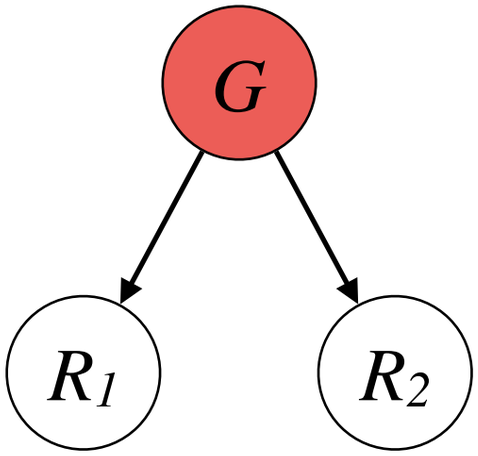


$$\text{data} = \{(?, 4, 5), (?, 4, 4), (?, 5, 3), (?, 1, 2), (?, 5, 4)\}$$

```
For each data point:
  "guess" the value of G
```

complete data

Expectation maximization



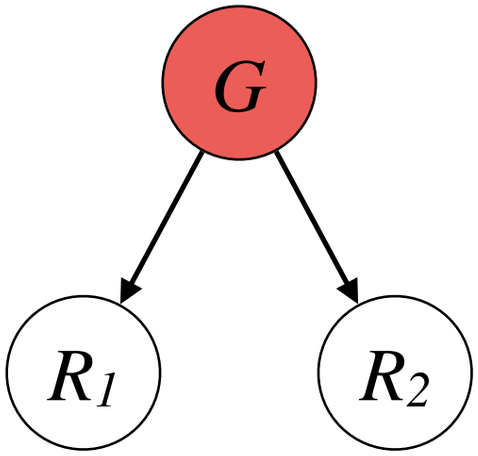
data = {(?, 4, 5), (?, 4, 4), (?, 5, 3), (?, 1, 2), (?, 5, 4)}

```
For each data point:
  "guess" the value of G
```

complete data

```
Learn parameters from
complete data.
```

Expectation maximization



data = {(?, 4, 5), (?, 4, 4), (?, 5, 3), (?, 1, 2), (?, 5, 4)}

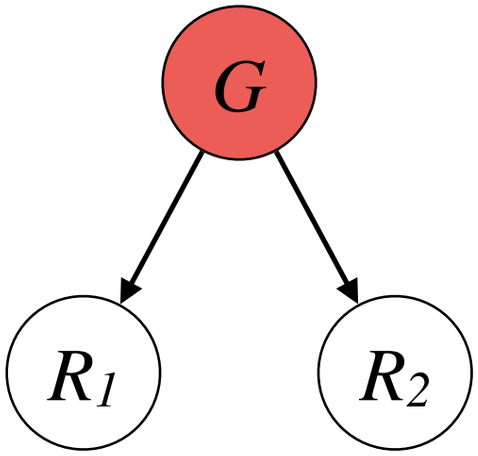
For each data point:
"guess" the value of G

complete data

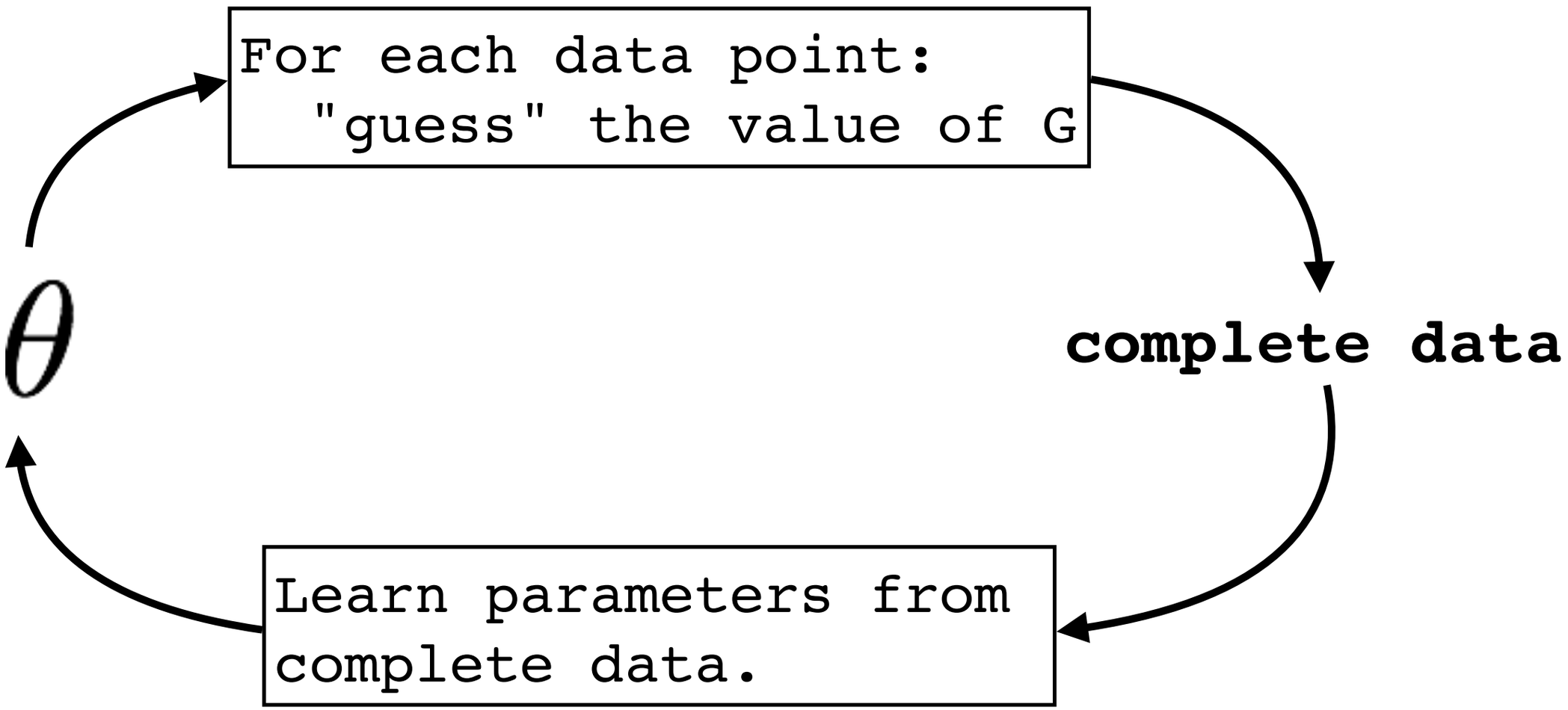
θ

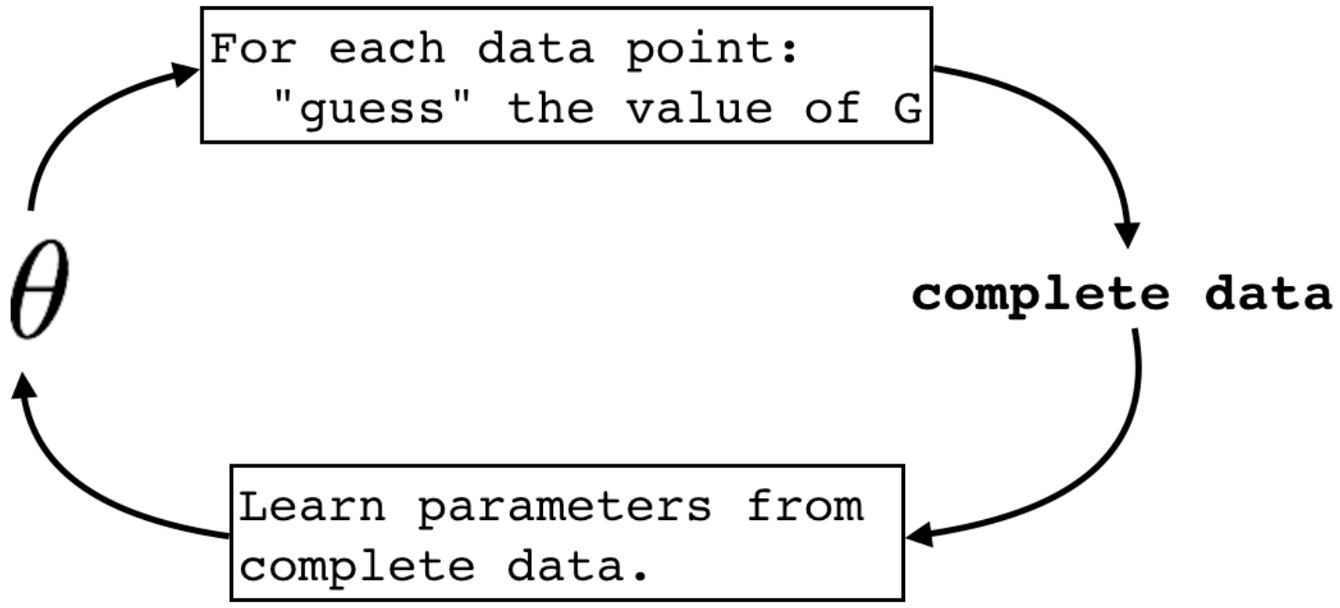
Learn parameters from complete data.

Expectation maximization

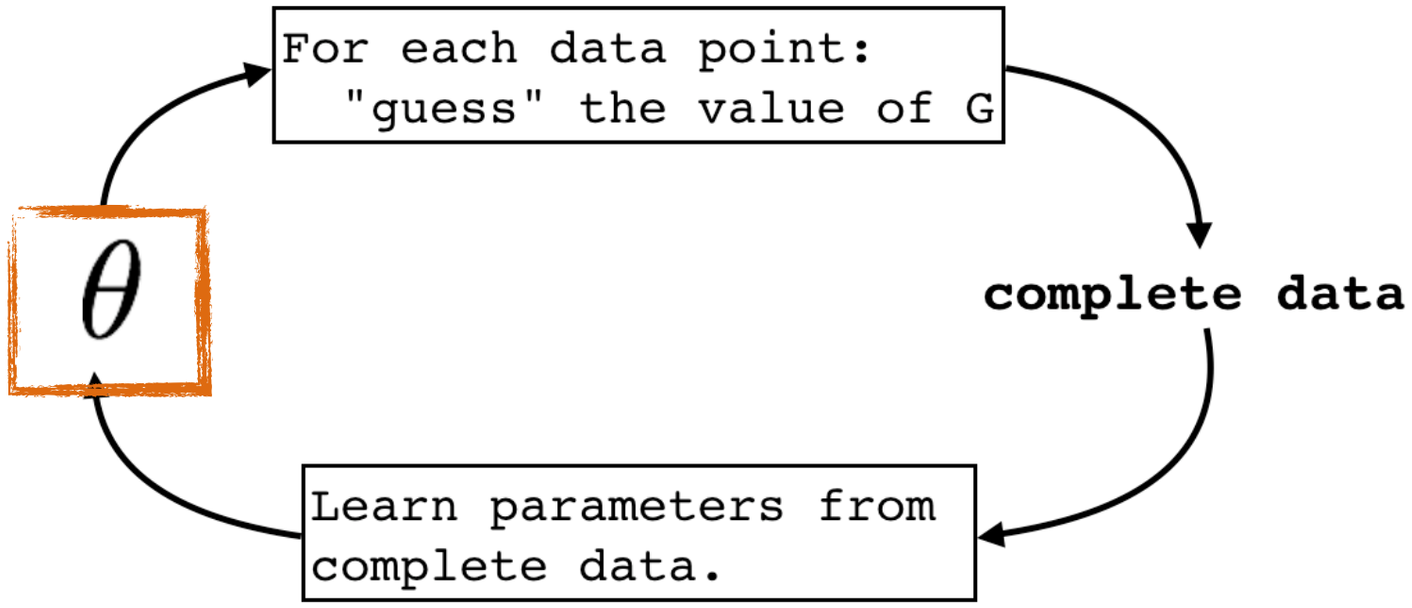


data = {(?, 4, 5), (?, 4, 4), (?, 5, 3), (?, 1, 2), (?, 5, 4)}

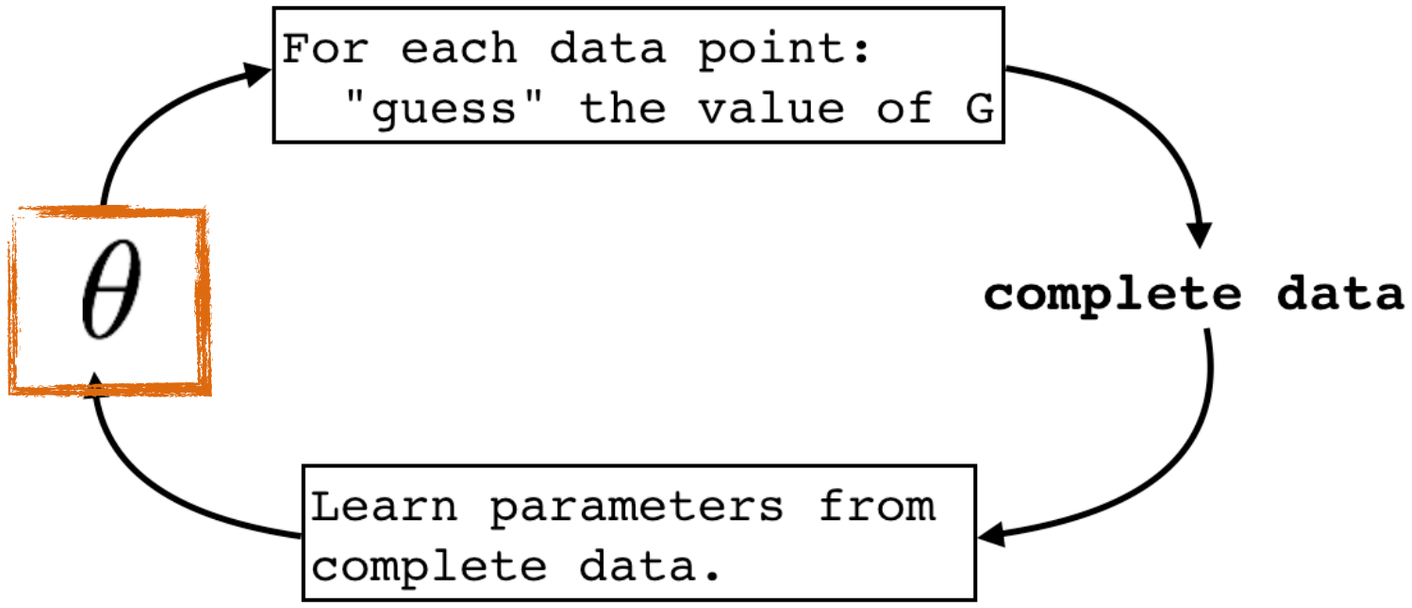




$$\text{data} = \{(?, 4, 5), (?, 4, 4), (?, 5, 3), (?, 1, 2), (?, 5, 4)\}$$



$$\text{data} = \{(?, 4, 5), (?, 4, 4), (?, 5, 3), (?, 1, 2), (?, 5, 4)\}$$



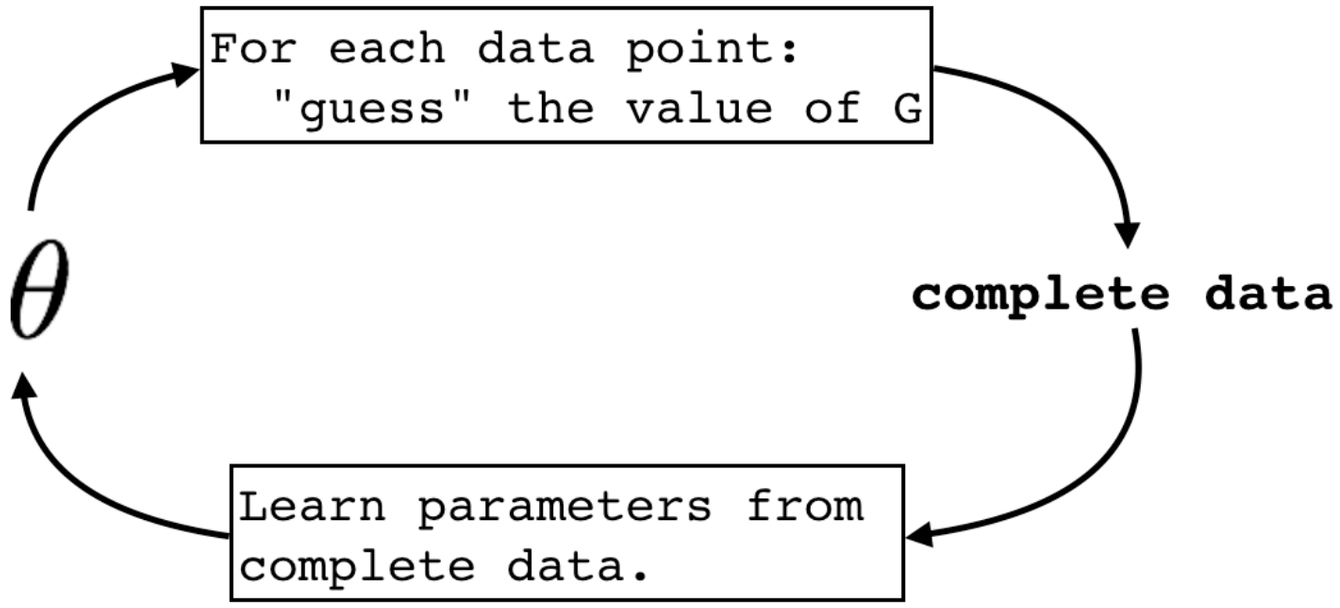
$$\text{data} = \{(?, 4, 5), (?, 4, 4), (?, 5, 3), (?, 1, 2), (?, 5, 4)\}$$

θ:

g	$p_G(g)$
d	3/5
c	2/5

g	r_1	$p_{R_1}(r \mid g)$
d	4	2/3
d	5	1/3
c	1	1/2
c	5	1/2

g	r_2	$p_{R_2}(r \mid g)$
d	3	1/3
d	4	1/3
d	5	1/3
c	2	1/2
c	4	1/2



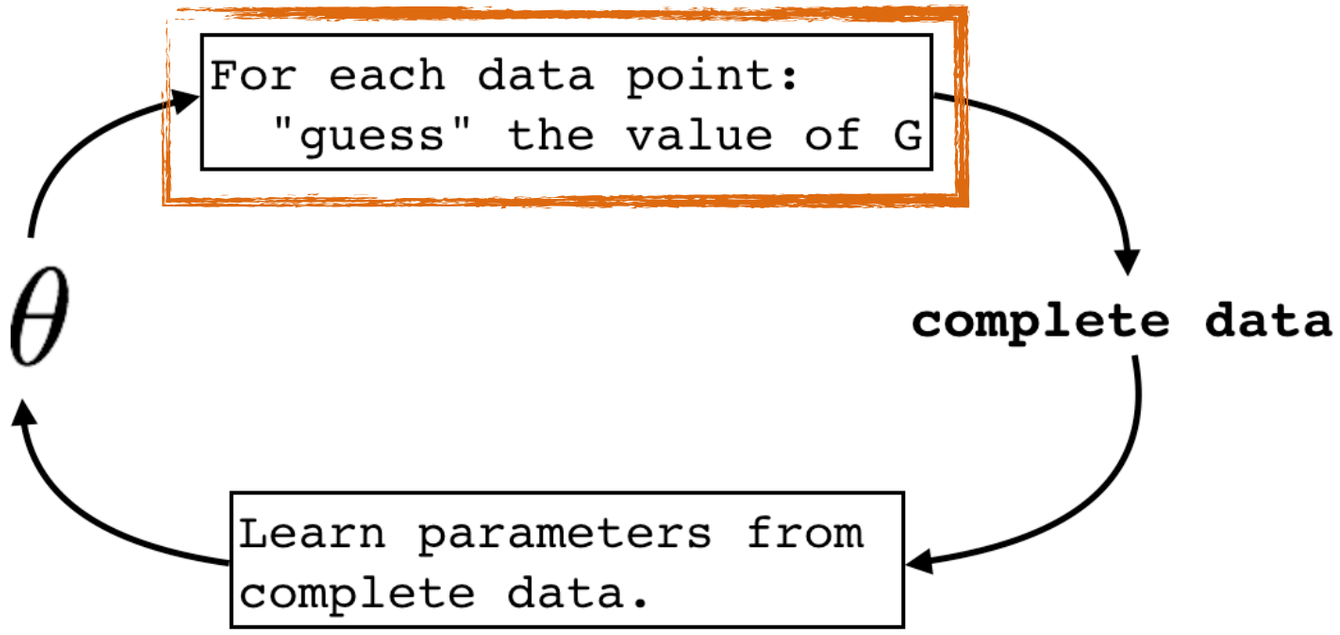
$$\text{data} = \{(?, 4, 5), (?, 4, 4), (?, 5, 3), (?, 1, 2), (?, 5, 4)\}$$

θ:

g	$p_G(g)$
d	3/5
c	2/5

g	r_1	$p_{R_1}(r \mid g)$
d	4	2/3
d	5	1/3
c	1	1/2
c	5	1/2

g	r_2	$p_{R_2}(r \mid g)$
d	3	1/3
d	4	1/3
d	5	1/3
c	2	1/2
c	4	1/2



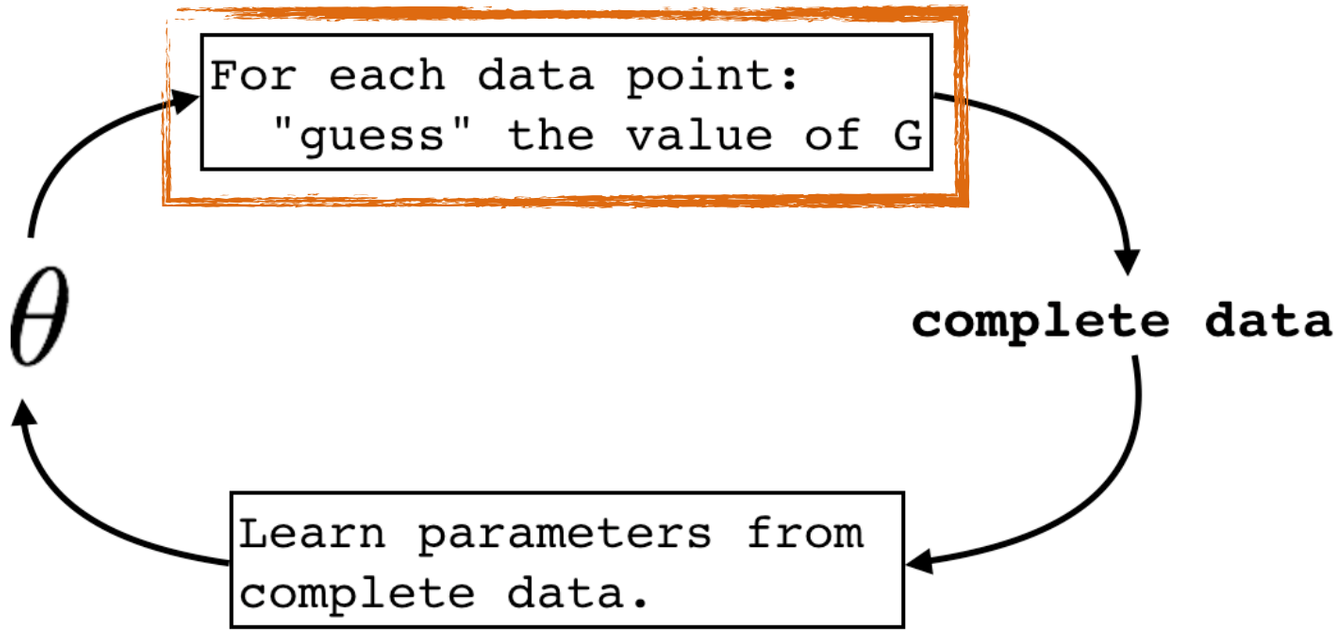
$$\text{data} = \{(?, 4, 5), (?, 4, 4), (?, 5, 3), (?, 1, 2), (?, 5, 4)\}$$

θ:

g	$p_G(g)$
d	3/5
c	2/5

g	r_1	$p_{R_1}(r \mid g)$
d	4	2/3
d	5	1/3
c	1	1/2
c	5	1/2

g	r_2	$p_{R_2}(r \mid g)$
d	3	1/3
d	4	1/3
d	5	1/3
c	2	1/2
c	4	1/2



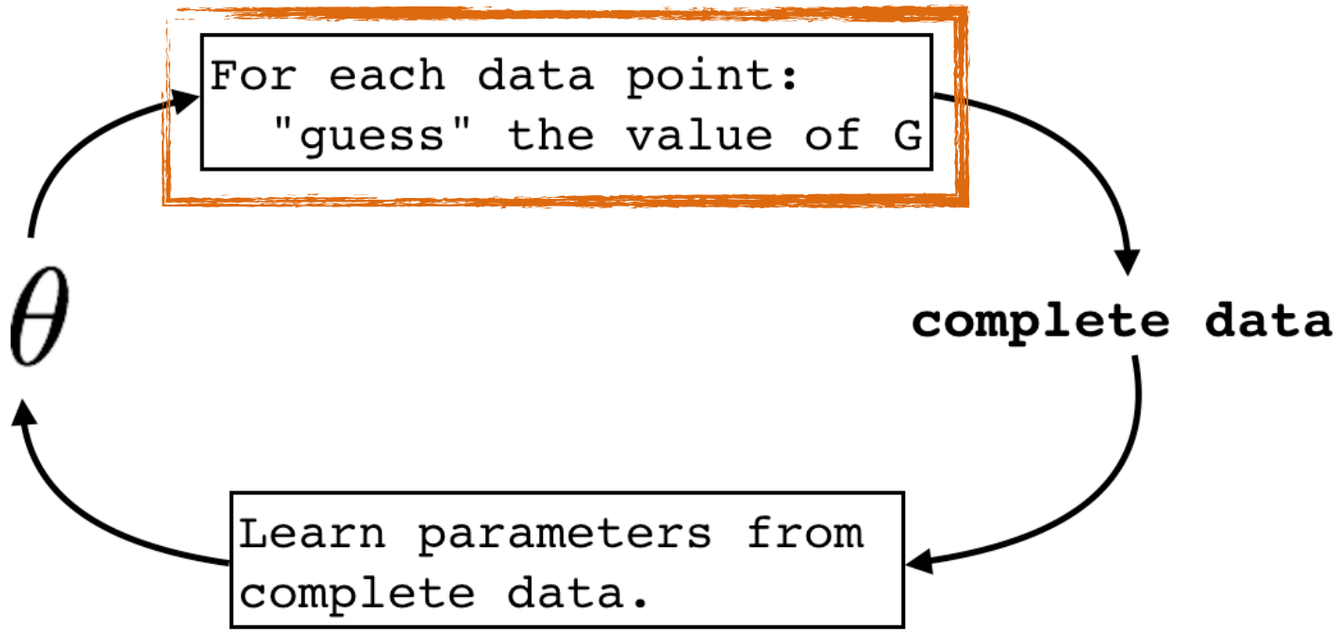
$$\text{data} = \{(?, 4, 5), (?, 4, 4), (?, 5, 3), (?, 1, 2), (?, 5, 4)\}$$

θ:

g	$p_G(g)$
d	3/5
c	2/5

g	r_1	$p_{R_1}(r \mid g)$
d	4	2/3
d	5	1/3
c	1	1/2
c	5	1/2

g	r_2	$p_{R_2}(r \mid g)$
d	3	1/3
d	4	1/3
d	5	1/3
c	2	1/2
c	4	1/2



$$\text{data} = \{(?, 4, 5), (?, 4, 4), (?, 5, 3), (?, 1, 2), (?, 5, 4)\}$$

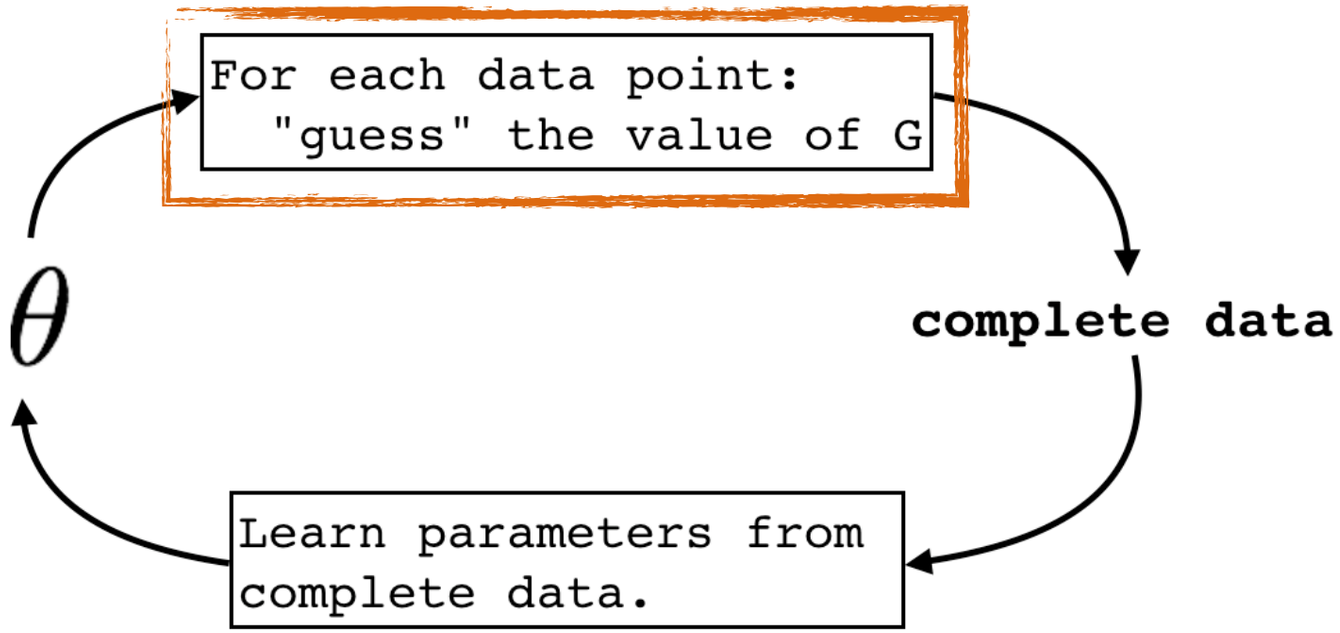
$$p(g \mid r_1{=}4, r_2{=}5)$$

θ:

g	$p_G(g)$
d	3/5
c	2/5

g	r_1	$p_{R_1}(r \mid g)$
d	4	2/3
d	5	1/3
c	1	1/2
c	5	1/2

g	r_2	$p_{R_2}(r \mid g)$
d	3	1/3
d	4	1/3
d	5	1/3
c	2	1/2
c	4	1/2



$$\text{data} = \{(?, 4, 5), (?, 4, 4), (?, 5, 3), (?, 1, 2), (?, 5, 4)\}$$

$$p(g \mid r_1{=}4, r_2{=}5)$$

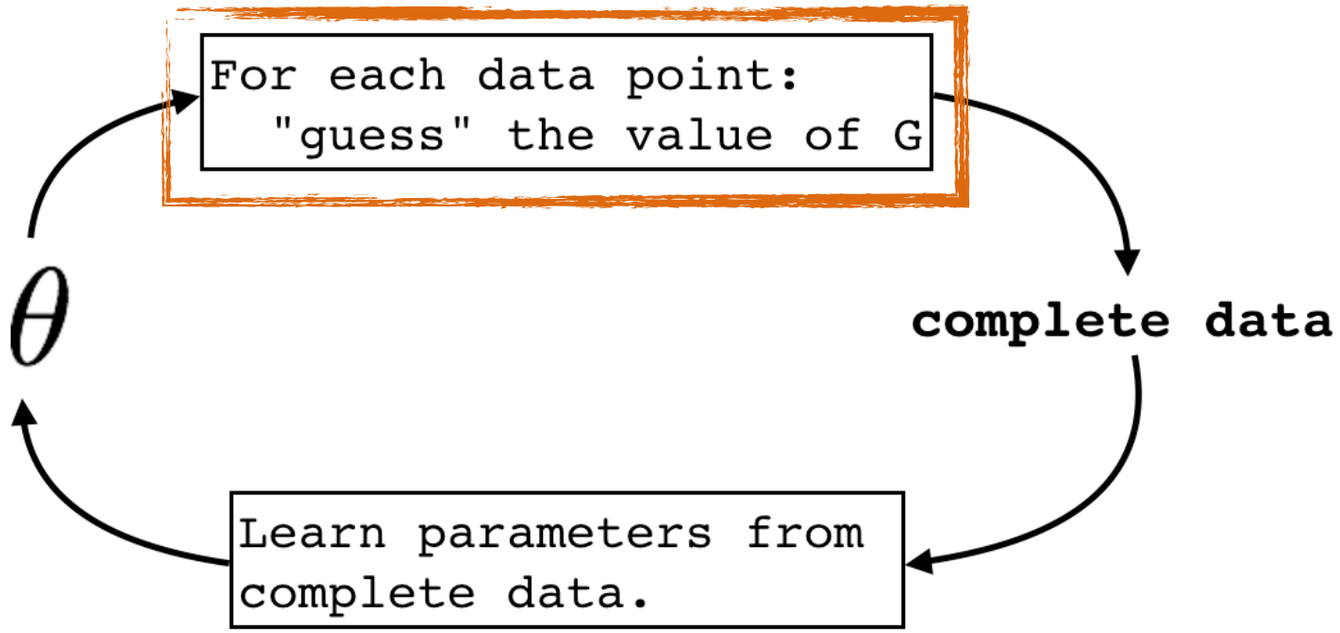
0.25 (**c**, 4, 5)

θ:

g	$p_G(g)$
d	3/5
c	2/5

g	r_1	$p_{R_1}(r \mid g)$
d	4	2/3
d	5	1/3
c	1	1/2
c	5	1/2

g	r_2	$p_{R_2}(r \mid g)$
d	3	1/3
d	4	1/3
d	5	1/3
c	2	1/2
c	4	1/2



$$\text{data} = \{(?, 4, 5), (?, 4, 4), (?, 5, 3), (?, 1, 2), (?, 5, 4)\}$$

$$p(g \mid r_1{=}4, r_2{=}5)$$

0.25 (**c**, 4, 5)

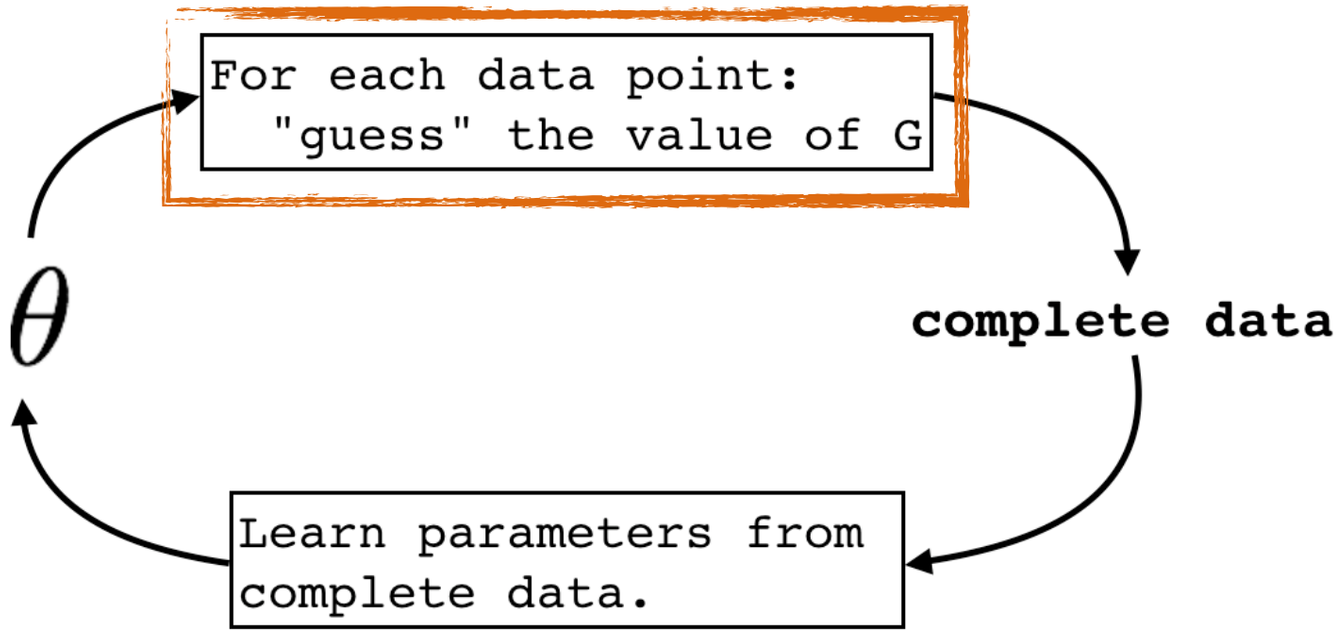
0.75 (**d**, 4, 5)

θ:

g	$p_G(g)$
d	3/5
c	2/5

g	r_1	$p_{R_1}(r \mid g)$
d	4	2/3
d	5	1/3
c	1	1/2
c	5	1/2

g	r_2	$p_{R_2}(r \mid g)$
d	3	1/3
d	4	1/3
d	5	1/3
c	2	1/2
c	4	1/2



$$\text{data} = \{(?, 4, 5), (?, 4, 4), (?, 5, 3), (?, 1, 2), (?, 5, 4)\}$$

0.25 (**c**, 4, 5)

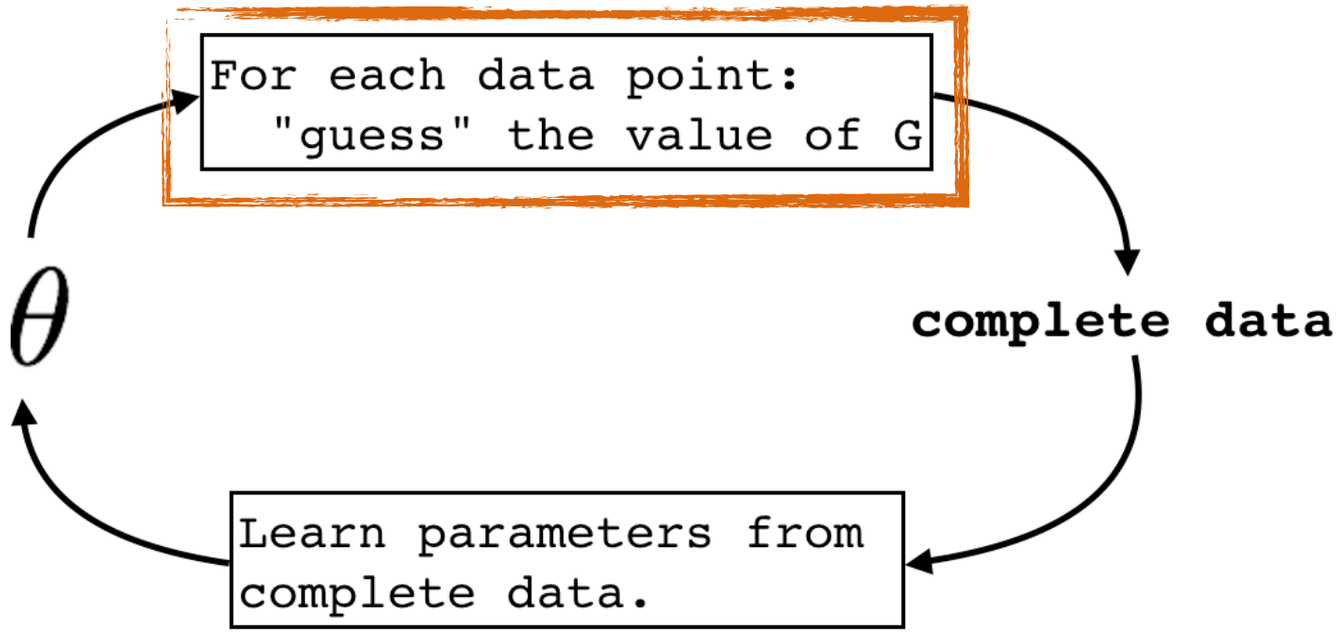
0.75 (**d**, 4, 5)

θ:

g	$p_G(g)$
d	3/5
c	2/5

g	r_1	$p_{R_1}(r \mid g)$
d	4	2/3
d	5	1/3
c	1	1/2
c	5	1/2

g	r_2	$p_{R_2}(r \mid g)$
d	3	1/3
d	4	1/3
d	5	1/3
c	2	1/2
c	4	1/2



$$\text{data} = \{(?, 4, 5), (?, 4, 4), (?, 5, 3), (?, 1, 2), (?, 5, 4)\}$$

$$p(g \mid r_1=4, r_2=4)$$

0.25 (**c**, 4, 5)

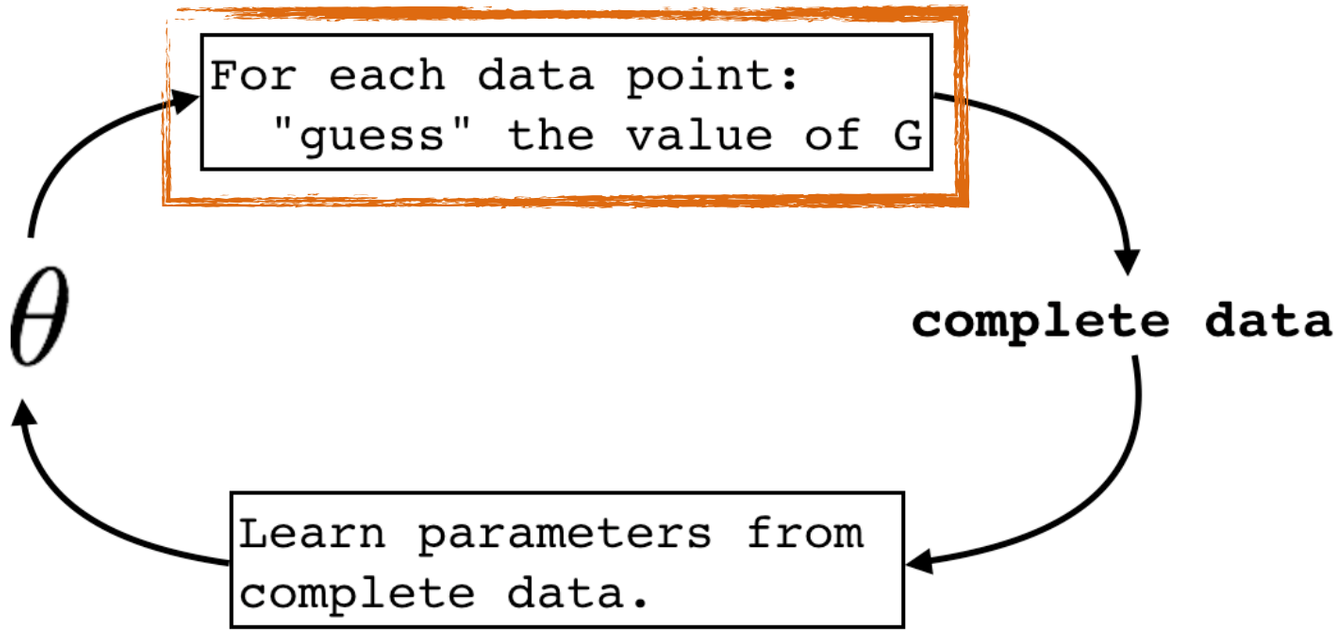
0.75 (**d**, 4, 5)

θ:

g	$p_G(g)$
d	3/5
c	2/5

g	r_1	$p_{R_1}(r \mid g)$
d	4	2/3
d	5	1/3
c	1	1/2
c	5	1/2

g	r_2	$p_{R_2}(r \mid g)$
d	3	1/3
d	4	1/3
d	5	1/3
c	2	1/2
c	4	1/2



$$\text{data} = \{(?, 4, 5), (?, 4, 4), (?, 5, 3), (?, 1, 2), (?, 5, 4)\}$$

$$p(g \mid r_1{=}4, r_2{=}4)$$

0.25 (**c**, 4, 5) **0.1** (**c**, 4, 4)

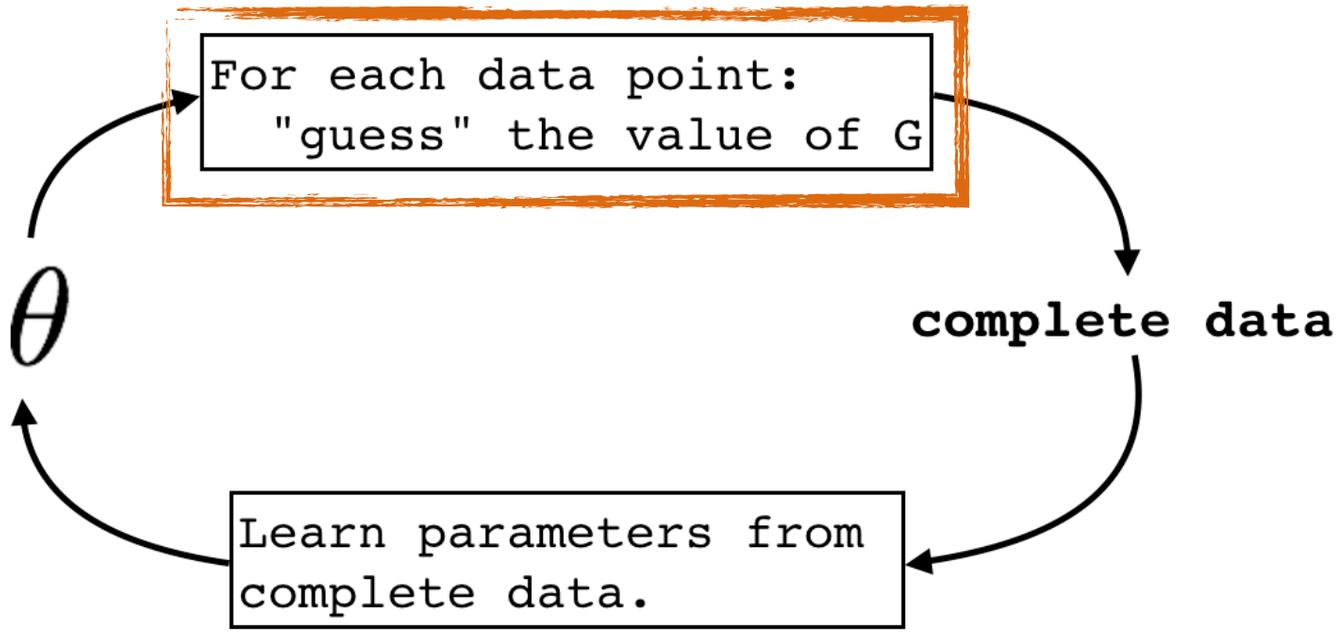
0.75 (**d**, 4, 5)

θ:

g	$p_G(g)$
d	3/5
c	2/5

g	r_1	$p_{R_1}(r \mid g)$
d	4	2/3
d	5	1/3
c	1	1/2
c	5	1/2

g	r_2	$p_{R_2}(r \mid g)$
d	3	1/3
d	4	1/3
d	5	1/3
c	2	1/2
c	4	1/2



$$\text{data} = \{(?, 4, 5), (?, 4, 4), (?, 5, 3), (?, 1, 2), (?, 5, 4)\}$$

$$p(g \mid r_1=4, r_2=4)$$

0.25 (**c**, 4, 5) **0.1** (**c**, 4, 4)

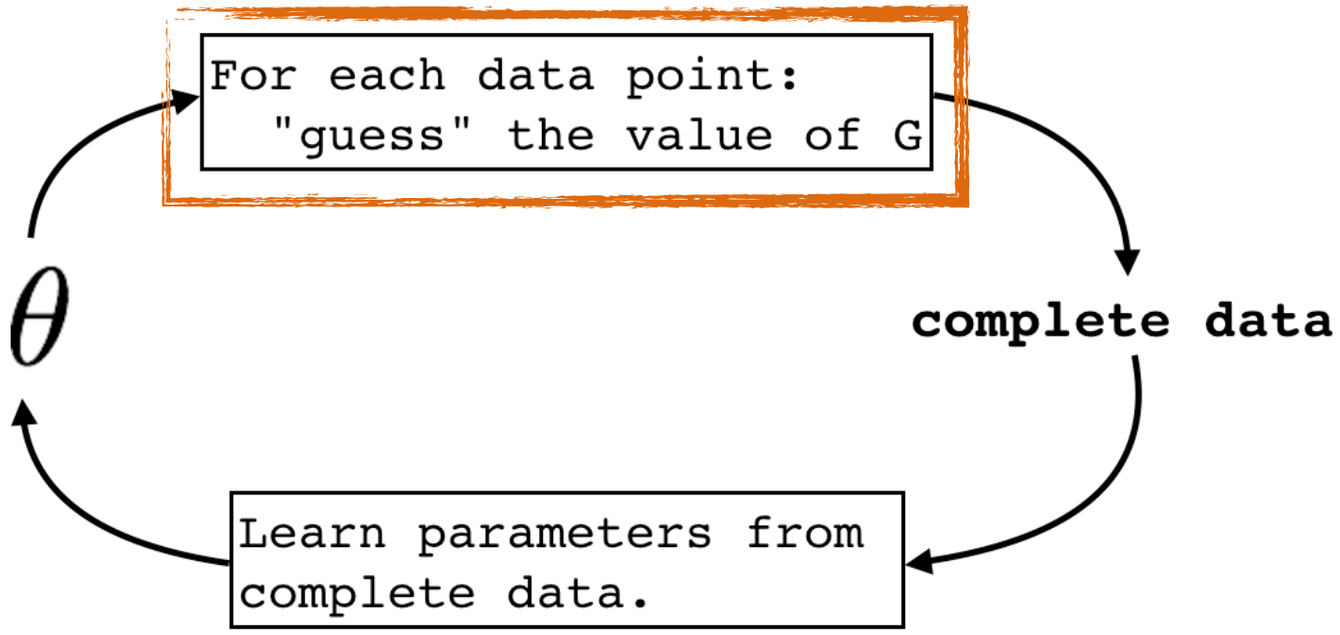
0.75 (**d**, 4, 5) **0.9** (**d**, 4, 4)

θ:

g	$p_G(g)$
d	3/5
c	2/5

g	r_1	$p_{R_1}(r \mid g)$
d	4	2/3
d	5	1/3
c	1	1/2
c	5	1/2

g	r_2	$p_{R_2}(r \mid g)$
d	3	1/3
d	4	1/3
d	5	1/3
c	2	1/2
c	4	1/2



$$\text{data} = \{(?, 4, 5), (?, 4, 4), (?, 5, 3), (?, 1, 2), (?, 5, 4)\}$$

0.25 (**c**, 4, 5) **0.1** (**c**, 4, 4)

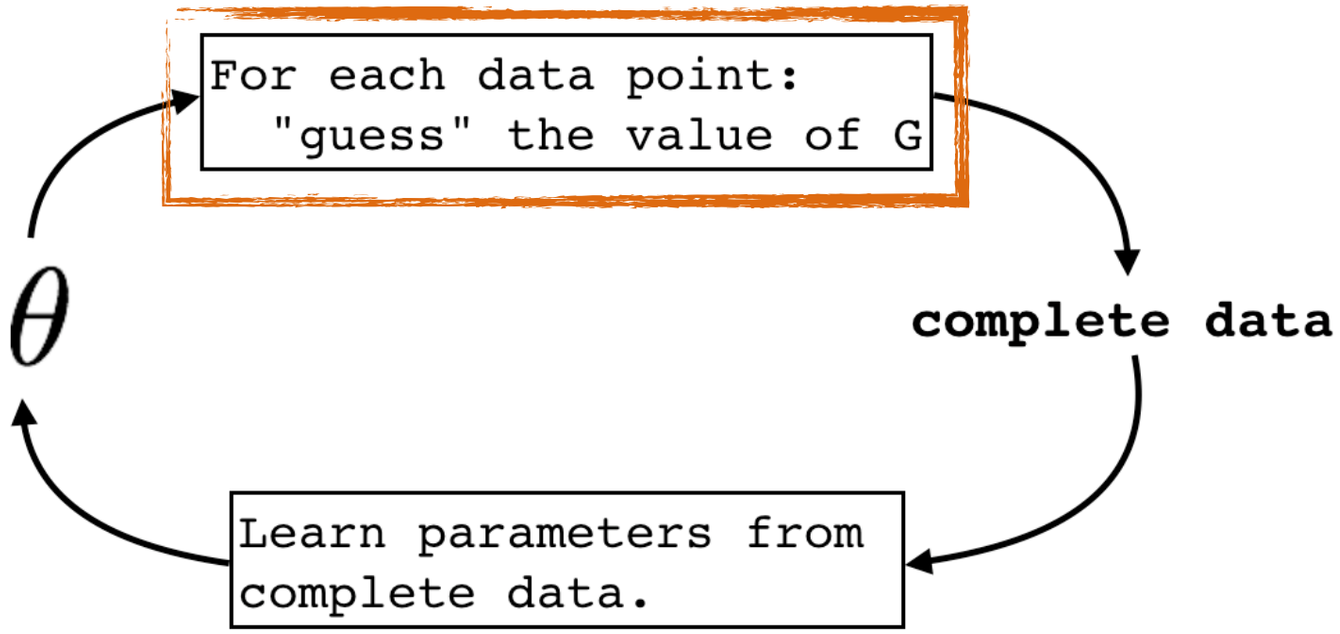
0.75 (**d**, 4, 5) **0.9** (**d**, 4, 4)

θ:

g	$p_G(g)$
d	3/5
c	2/5

g	r_1	$p_{R_1}(r \mid g)$
d	4	2/3
d	5	1/3
c	1	1/2
c	5	1/2

g	r_2	$p_{R_2}(r \mid g)$
d	3	1/3
d	4	1/3
d	5	1/3
c	2	1/2
c	4	1/2



$$\text{data} = \{(?, 4, 5), (?, 4, 4), (?, 5, 3), (?, 1, 2), (?, 5, 4)\}$$

$$p(g \mid r_1{=}5, r_2{=}3)$$

0.25 (**c**, 4, 5) **0.1** (**c**, 4, 4)

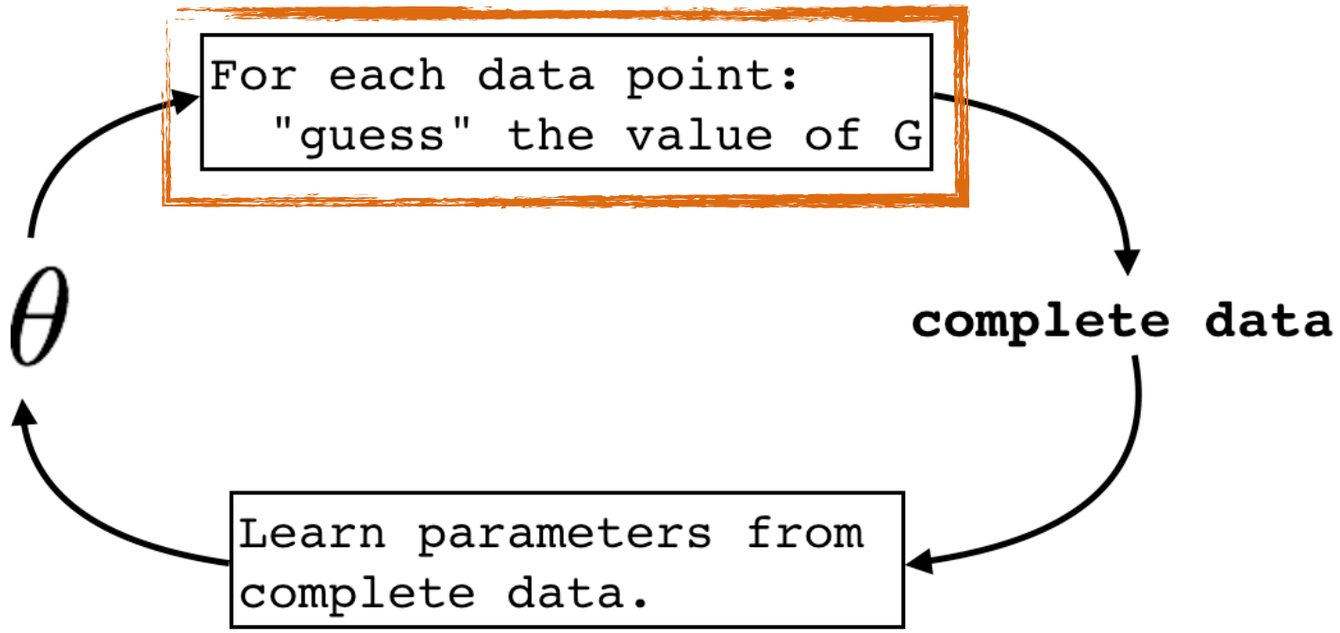
0.75 (**d**, 4, 5) **0.9** (**d**, 4, 4)

θ:

g	$p_G(g)$
d	3/5
c	2/5

g	r_1	$p_{R_1}(r \mid g)$
d	4	2/3
d	5	1/3
c	1	1/2
c	5	1/2

g	r_2	$p_{R_2}(r \mid g)$
d	3	1/3
d	4	1/3
d	5	1/3
c	2	1/2
c	4	1/2



$$\text{data} = \{(?, 4, 5), (?, 4, 4), (?, 5, 3), (?, 1, 2), (?, 5, 4)\}$$

$$p(g \mid r_1{=}5, r_2{=}3)$$

0.25 (**c**, 4, 5) **0.1** (**c**, 4, 4) **0.8** (**c**, 5, 3)

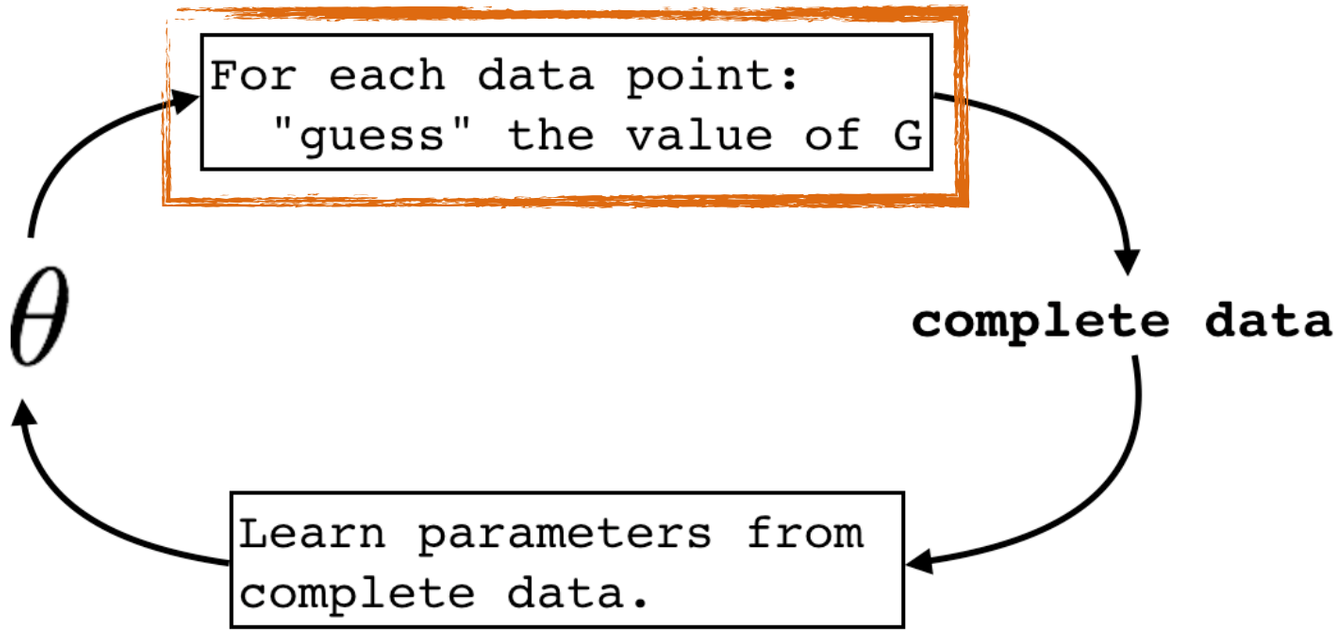
0.75 (**d**, 4, 5) **0.9** (**d**, 4, 4) **0.2** (**d**, 5, 3)

θ:

g	$p_G(g)$
d	3/5
c	2/5

g	r_1	$p_{R_1}(r \mid g)$
d	4	2/3
d	5	1/3
c	1	1/2
c	5	1/2

g	r_2	$p_{R_2}(r \mid g)$
d	3	1/3
d	4	1/3
d	5	1/3
c	2	1/2
c	4	1/2



$$\text{data} = \{(?, 4, 5), (?, 4, 4), (?, 5, 3), (?, 1, 2), (?, 5, 4)\}$$

0.25 (**c**, 4, 5) **0.1** (**c**, 4, 4) **0.8** (**c**, 5, 3) **0.9** (**c**, 1, 2) **0.7** (**c**, 5, 4)

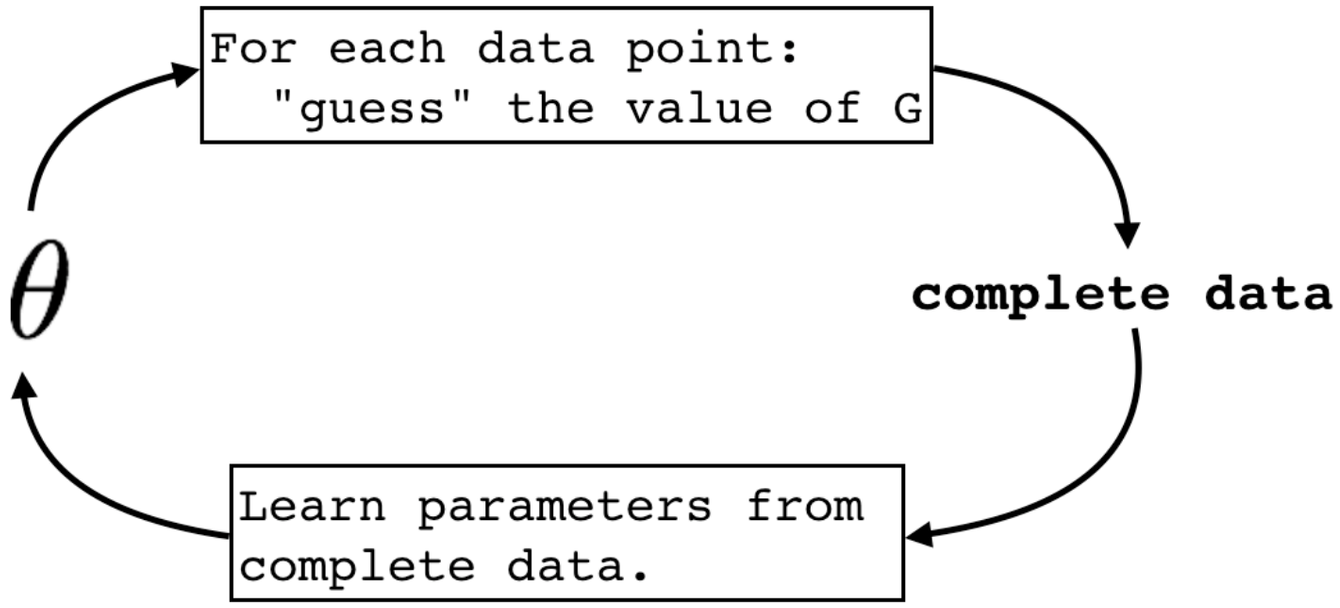
0.75 (**d**, 4, 5) **0.9** (**d**, 4, 4) **0.2** (**d**, 5, 3) **0.1** (**d**, 1, 2) **0.3** (**d**, 5, 4)

θ:

g	$p_G(g)$
d	3/5
c	2/5

g	r_1	$p_{R_1}(r \mid g)$
d	4	2/3
d	5	1/3
c	1	1/2
c	5	1/2

g	r_2	$p_{R_2}(r \mid g)$
d	3	1/3
d	4	1/3
d	5	1/3
c	2	1/2
c	4	1/2



$$\text{data} = \{(?, 4, 5), (?, 4, 4), (?, 5, 3), (?, 1, 2), (?, 5, 4)\}$$

0.25 (**c**, 4, 5) **0.1** (**c**, 4, 4) **0.8** (**c**, 5, 3) **0.9** (**c**, 1, 2) **0.7** (**c**, 5, 4)

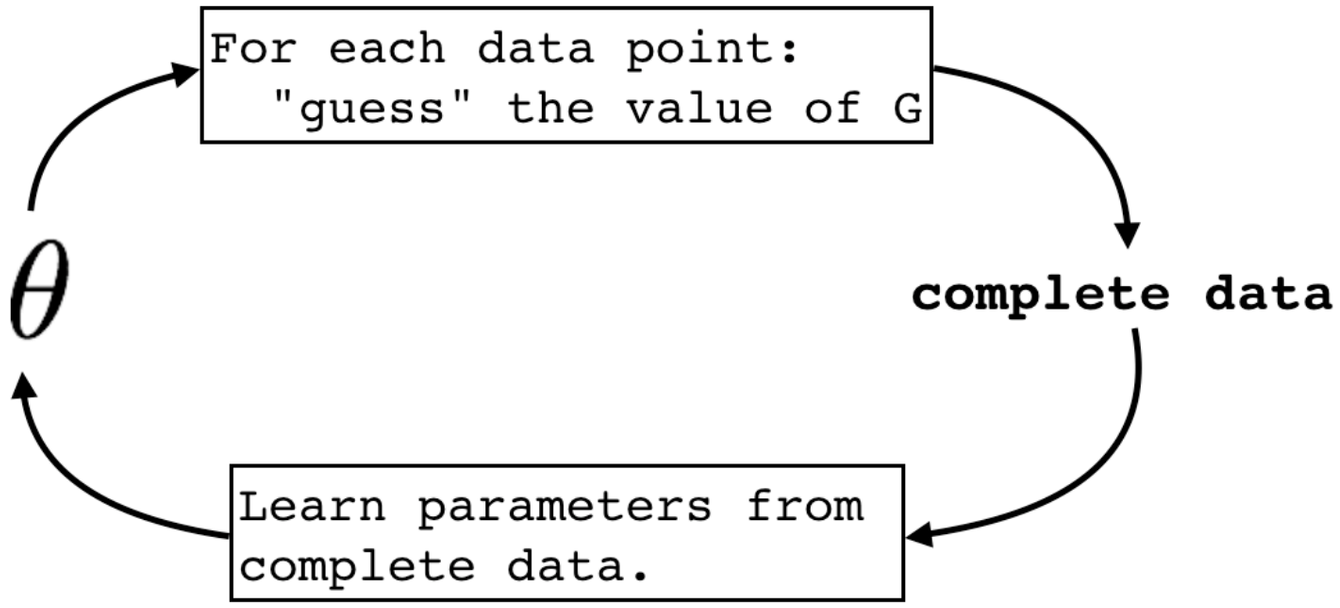
0.75 (**d**, 4, 5) **0.9** (**d**, 4, 4) **0.2** (**d**, 5, 3) **0.1** (**d**, 1, 2) **0.3** (**d**, 5, 4)

θ:

g	$p_G(g)$
d	3/5
c	2/5

g	r_1	$p_{R_1}(r \mid g)$
d	4	2/3
d	5	1/3
c	1	1/2
c	5	1/2

g	r_2	$p_{R_2}(r \mid g)$
d	3	1/3
d	4	1/3
d	5	1/3
c	2	1/2
c	4	1/2



$$\text{data} = \{(?, 4, 5), (?, 4, 4), (?, 5, 3), (?, 1, 2), (?, 5, 4)\}$$

0.25 (**c**, 4, 5) **0.1** (**c**, 4, 4) **0.8** (**c**, 5, 3) **0.9** (**c**, 1, 2) **0.7** (**c**, 5, 4)

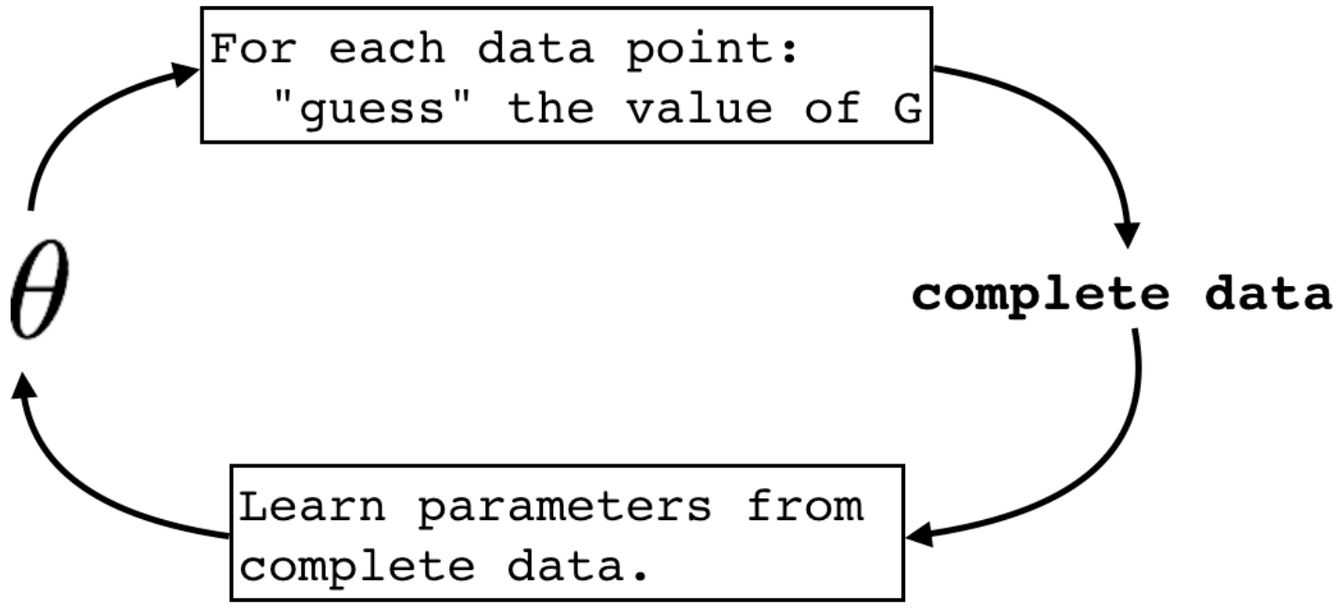
0.75 (**d**, 4, 5) **0.9** (**d**, 4, 4) **0.2** (**d**, 5, 3) **0.1** (**d**, 1, 2) **0.3** (**d**, 5, 4)

θ:

g	$p_G(g)$
d	3/5
c	2/5

g	r_1	$p_{R_1}(r \mid g)$
d	4	2/3
d	5	1/3
c	1	1/2
c	5	1/2

g	r_2	$p_{R_2}(r \mid g)$
d	3	1/3
d	4	1/3
d	5	1/3
c	2	1/2
c	4	1/2



$$\text{data} = \{(?, 4, 5), (?, 4, 4), (?, 5, 3), (?, 1, 2), (?, 5, 4)\}$$

complete data (weighted)

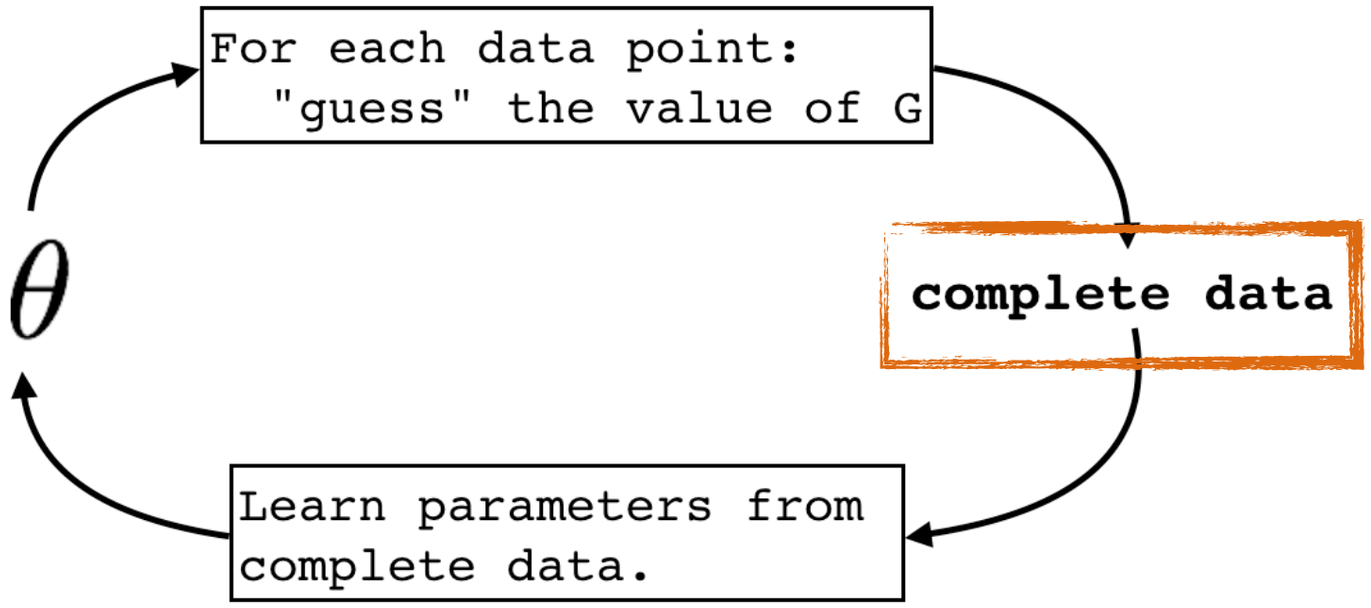
0.25 (**c**, 4, 5)	**0.1** (**c**, 4, 4)	**0.8** (**c**, 5, 3)	**0.9** (**c**, 1, 2)	**0.7** (**c**, 5, 4)
0.75 (**d**, 4, 5)	**0.9** (**d**, 4, 4)	**0.2** (**d**, 5, 3)	**0.1** (**d**, 1, 2)	**0.3** (**d**, 5, 4)

θ:

g	$p_G(g)$
d	3/5
c	2/5

g	r_1	$p_{R_1}(r \mid g)$
d	4	2/3
d	5	1/3
c	1	1/2
c	5	1/2

g	r_2	$p_{R_2}(r \mid g)$
d	3	1/3
d	4	1/3
d	5	1/3
c	2	1/2
c	4	1/2



$$\text{data} = \{(?, 4, 5), (?, 4, 4), (?, 5, 3), (?, 1, 2), (?, 5, 4)\}$$

complete data (weighted)

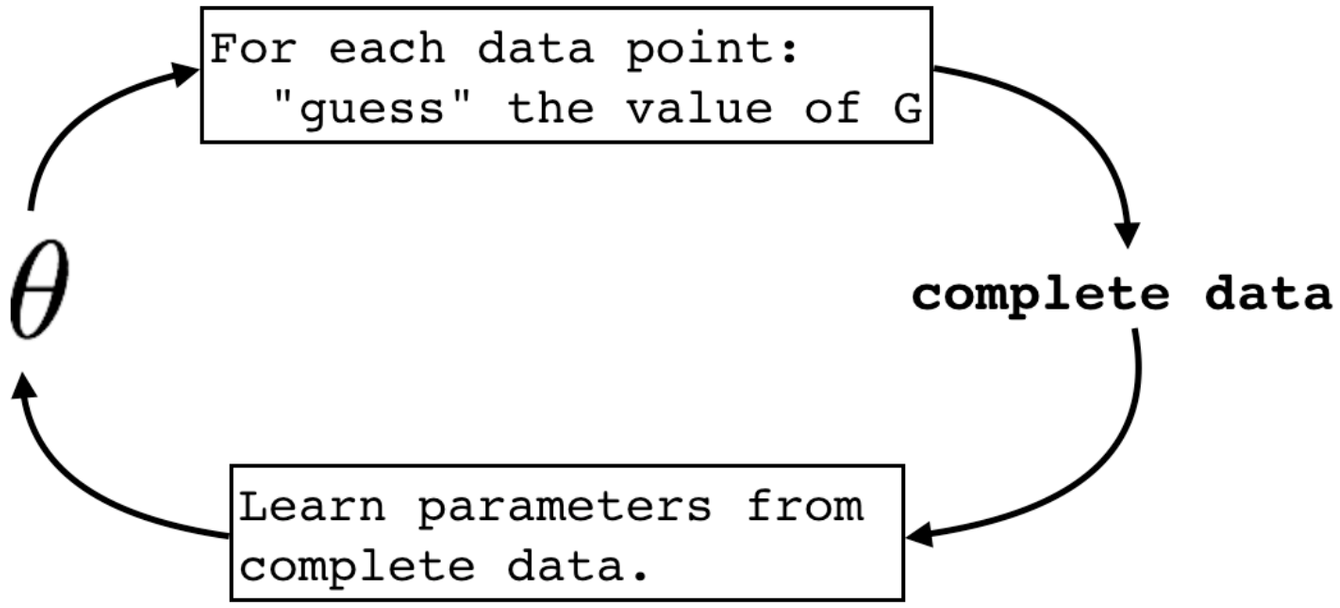
0.25 (**c**, 4, 5)	**0.1** (**c**, 4, 4)	**0.8** (**c**, 5, 3)	**0.9** (**c**, 1, 2)	**0.7** (**c**, 5, 4)
0.75 (**d**, 4, 5)	**0.9** (**d**, 4, 4)	**0.2** (**d**, 5, 3)	**0.1** (**d**, 1, 2)	**0.3** (**d**, 5, 4)

θ:

g	$p_G(g)$
d	3/5
c	2/5

g	r_1	$p_{R_1}(r \mid g)$
d	4	2/3
d	5	1/3
c	1	1/2
c	5	1/2

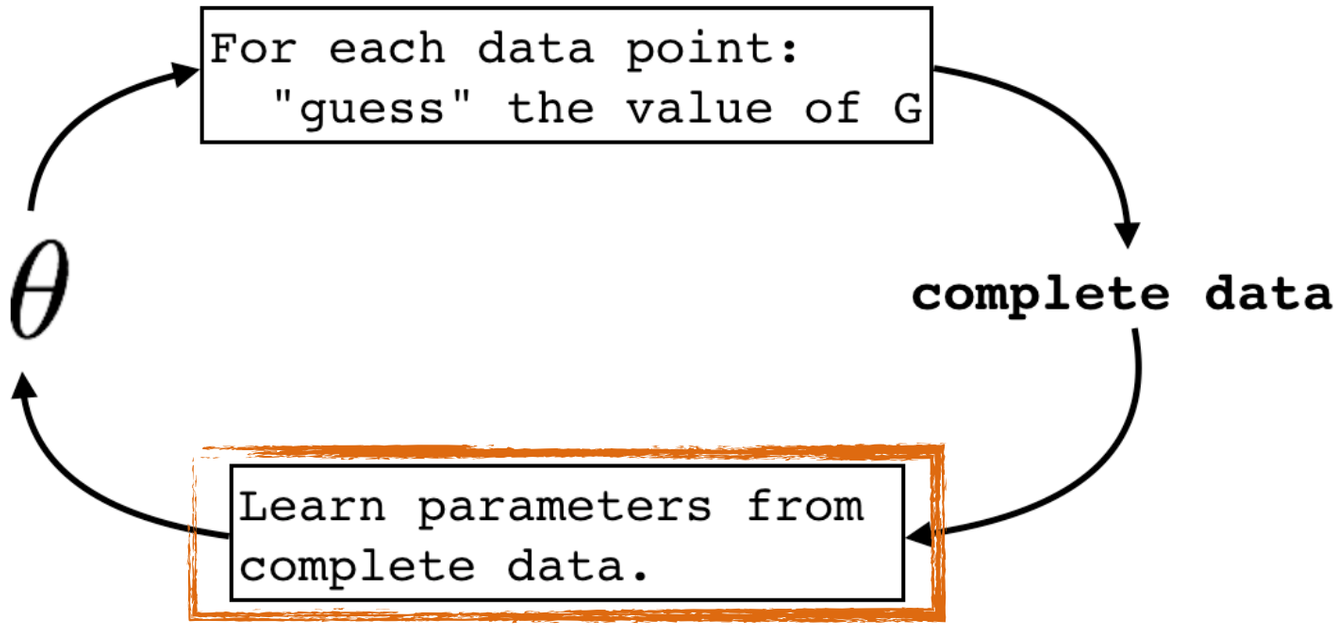
g	r_2	$p_{R_2}(r \mid g)$
d	3	1/3
d	4	1/3
d	5	1/3
c	2	1/2
c	4	1/2



$$\text{data} = \{(?, 4, 5), (?, 4, 4), (?, 5, 3), (?, 1, 2), (?, 5, 4)\}$$

complete data (weighted)

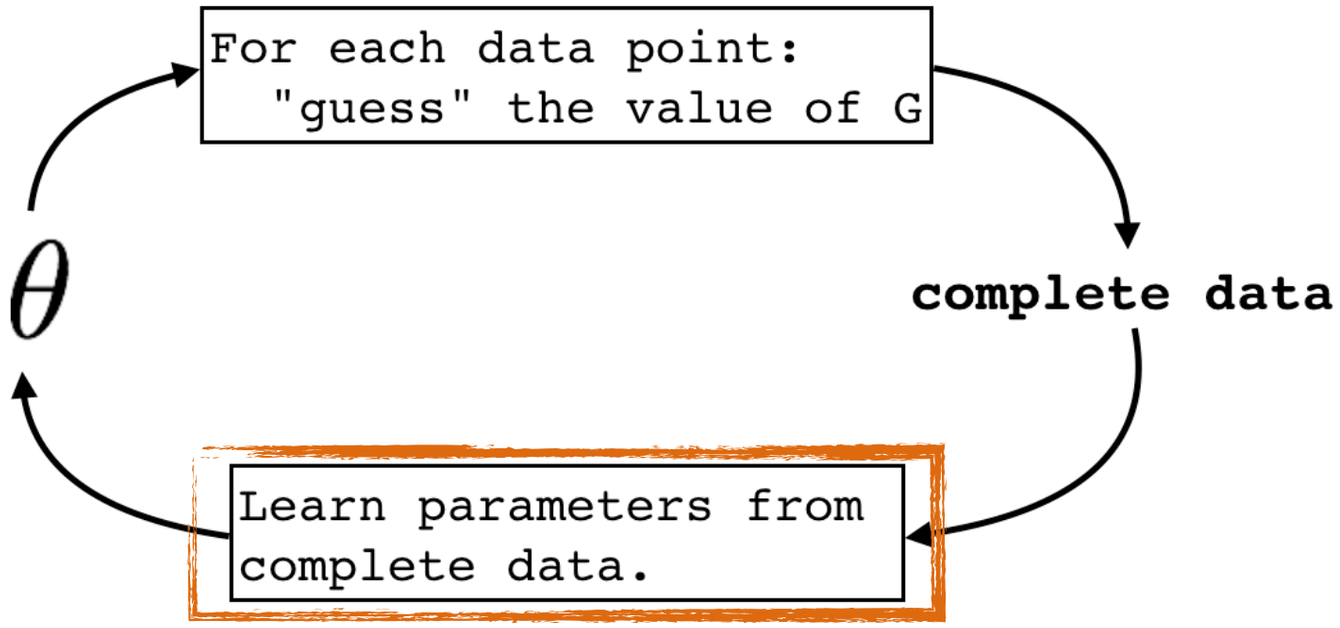
0.25 (**c**, 4, 5)	**0.1** (**c**, 4, 4)	**0.8** (**c**, 5, 3)	**0.9** (**c**, 1, 2)	**0.7** (**c**, 5, 4)
0.75 (**d**, 4, 5)	**0.9** (**d**, 4, 4)	**0.2** (**d**, 5, 3)	**0.1** (**d**, 1, 2)	**0.3** (**d**, 5, 4)



$$\text{data} = \{(?, 4, 5), (?, 4, 4), (?, 5, 3), (?, 1, 2), (?, 5, 4)\}$$

complete data (weighted)

0.25 (**c**, 4, 5)	**0.1** (**c**, 4, 4)	**0.8** (**c**, 5, 3)	**0.9** (**c**, 1, 2)	**0.7** (**c**, 5, 4)
0.75 (**d**, 4, 5)	**0.9** (**d**, 4, 4)	**0.2** (**d**, 5, 3)	**0.1** (**d**, 1, 2)	**0.3** (**d**, 5, 4)



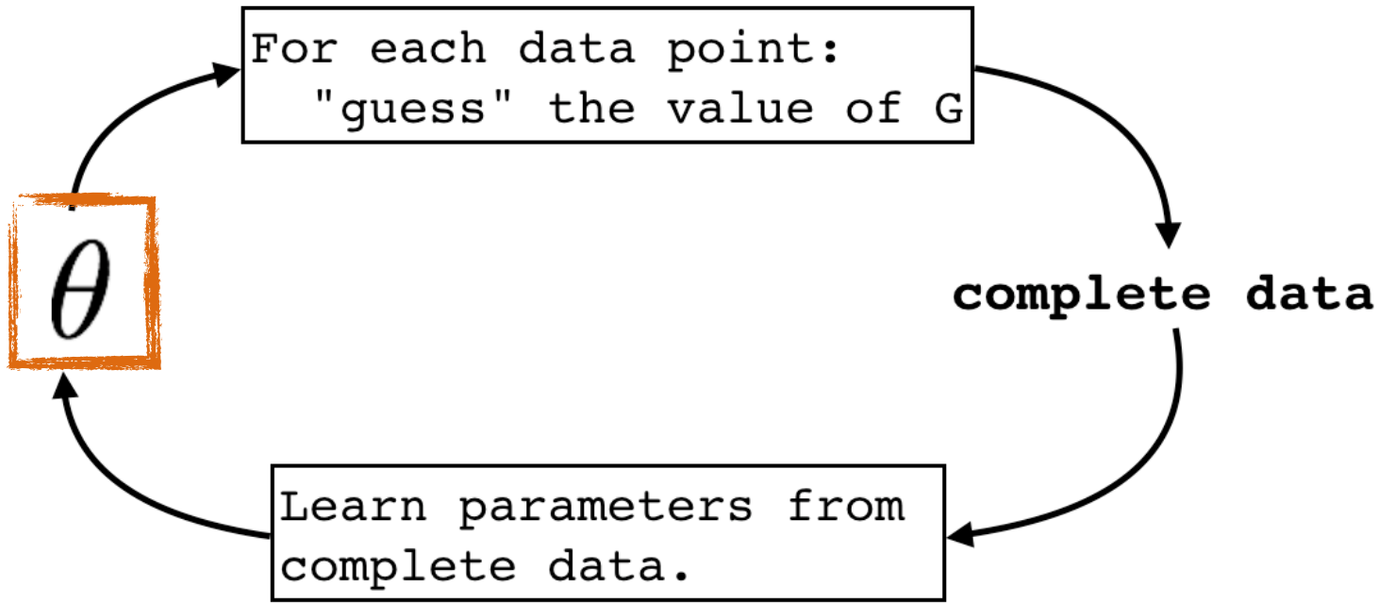
$$\text{data} = \{(?, 4, 5), (?, 4, 4), (?, 5, 3), (?, 1, 2), (?, 5, 4)\}$$

complete data (weighted)

0.25 (**c**, 4, 5)	**0.1** (**c**, 4, 4)	**0.8** (**c**, 5, 3)	**0.9** (**c**, 1, 2)	**0.7** (**c**, 5, 4)
0.75 (**d**, 4, 5)	**0.9** (**d**, 4, 4)	**0.2** (**d**, 5, 3)	**0.1** (**d**, 1, 2)	**0.3** (**d**, 5, 4)

count and normalize

θ



$$\text{data} = \{(?, 4, 5), (?, 4, 4), (?, 5, 3), (?, 1, 2), (?, 5, 4)\}$$

complete data (weighted)

0.25 (**c**, 4, 5)	**0.1** (**c**, 4, 4)	**0.8** (**c**, 5, 3)	**0.9** (**c**, 1, 2)	**0.7** (**c**, 5, 4)
0.75 (**d**, 4, 5)	**0.9** (**d**, 4, 4)	**0.2** (**d**, 5, 3)	**0.1** (**d**, 1, 2)	**0.3** (**d**, 5, 4)

count and normalize

θ