Basic Probability

Robert Platt
Northeastern University

Some images and slides are used from:
1. AIMA
2. Chris Amato

(Discrete) Random variables

What is a random variable?

Suppose that the variable *a* denotes the outcome of a role of a single six-sided die:

$$a \in \underbrace{\{1, 2, 3, 4, 5, 6\}} = A$$

a is a *random variable*

this is the *domain* of *a*

Another example:

Suppose *b* denotes whether it is raining or clear outside:

$$b \in \{rain, clear\} = B$$

Probability distribution

A probability distribution associates each with a probability of occurrence, represented by a *probability mass function (pmf)*.

A probability table is one way to encode the distribution:

$$a \in \{1, 2, 3, 4, 5, 6\} = A \qquad b \in \{rain, clear\} = B$$

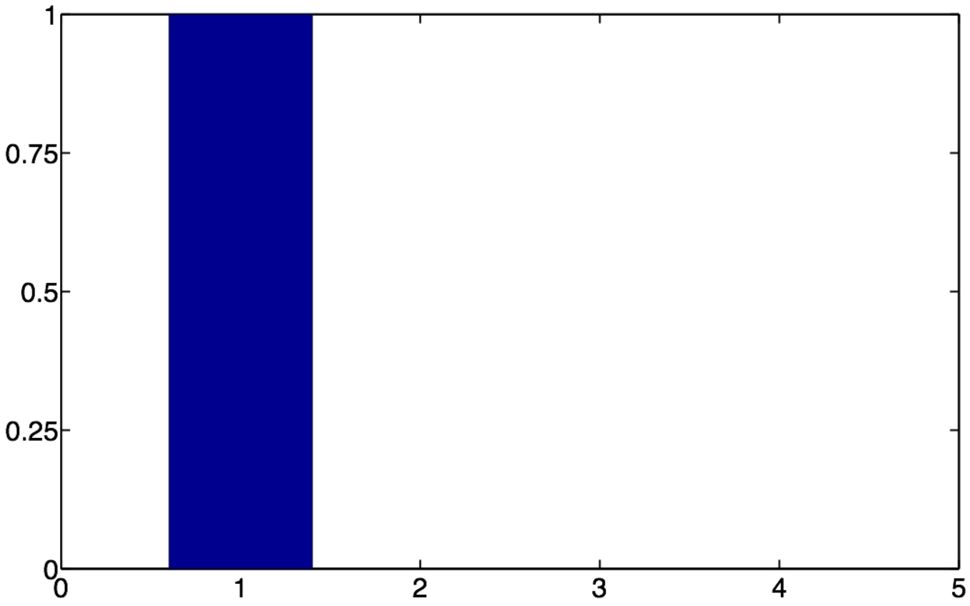
a	P(a)
1	1/6
2	1/6
3	1/6
4	1/6
5	1/6
6	1/6

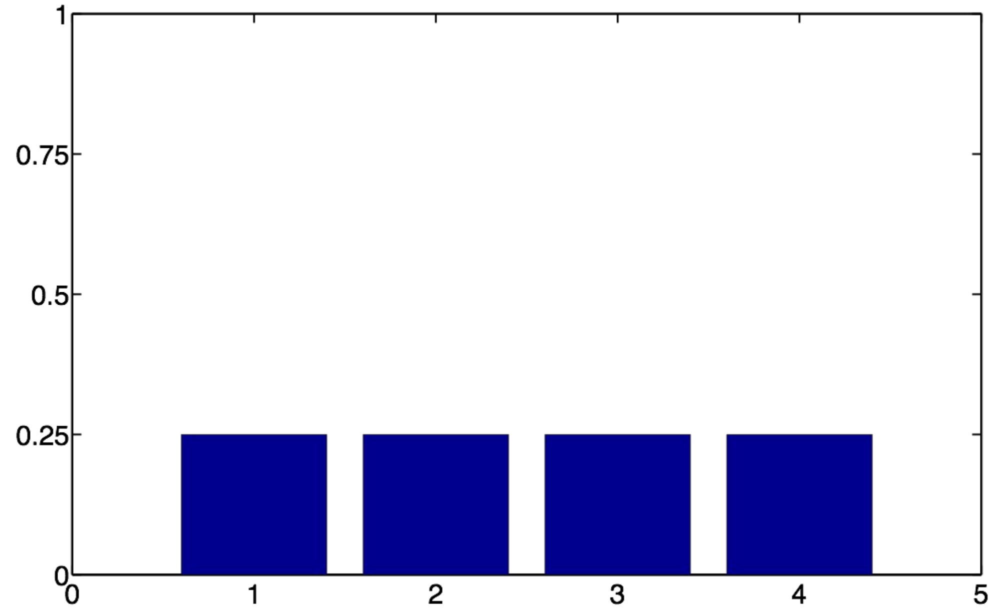
b	P(b)
rain	1/4
clear	3/4

<u>All probability distributions must satisfy the following:</u>

1. $\forall a \in A, a \geq 0$
2. $\sum_{a \in A} a = 1$

Example pmfs





Two pmfs over a state space of $X=\{1,2,3,4\}$

Writing probabilities

a	P(a)
1	1/6
2	1/6
3	1/6
4	1/6
5	1/6
6	1/6

b	P(b)
rain	1/4
clear	3/4

For example: $p(a = 2) = 1/6$

$$p(b = clear) = 3/4$$

But, sometimes we will abbreviate this as: $p(2) = 1/6$

$$p(clear) = 3/4$$

Types of random variables

Propositional or Boolean random variables

- e.g., *Cavity* (do I have a cavity?)
- *Cavity* = *true* is a proposition, also written *cavity*

Discrete random variables (finite or infinite)

- e.g., *Weather* is one of ⟨*sunny, rain, cloudy, snow*⟩
- *Weather* = *rain* is a proposition
- Values must be exhaustive and mutually exclusive

Continuous random variables (bounded or unbounded)

- e.g., Temp < 22.0

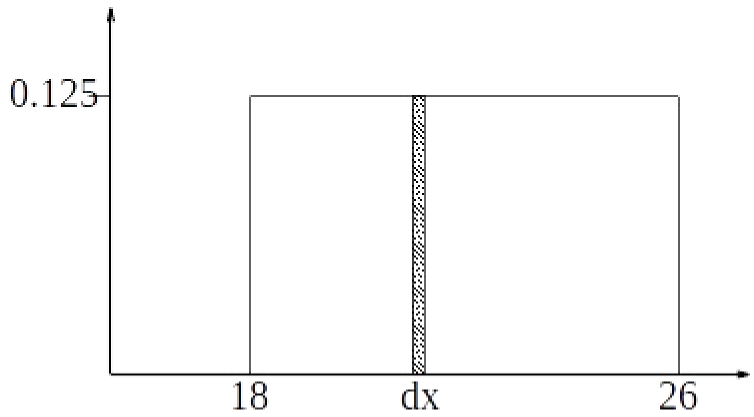
Continuous random variables

Cumulate distribution function (*cdf*), *F*(*q*)=(*X*<*q*) with *P*(*a*<*X*≤*b*)=*F*(*b*)-*F*(*a*)

Probability density function (*pdf*), $f(x) = \frac{d}{dx} F(x)$ with $P(a < X \leq b) = \int_a^b f(x)$

Express distribution as a parameterized function of value:

- e.g., *P*(*X* = *x*) = *U*[18, 26](*x*) = uniform density between 18 and 26



Here *P* is a density; integrates to 1.

P(*X* = 20.5) = 0.125 really means $\lim_{dx \to 0} P(20.5 \leq X \leq 20.5 + dx) / dx = 0.125$

Joint probability distributions

Given random variables: $X_1, X_2, \ldots, X_n$

The *joint distribution* is a probability assignment to all combinations: $P(X_1 = x_1, X_2 = x_2, \ldots, X_n = x_n)$

or: $P(x_1, x_2, \ldots, x_n)$

Sometimes written as: $P(X_1 = x_1 \wedge X_2 = x_2 \wedge \ldots \wedge X_n = x_n)$

As with single-variate distributions, joint distributions must satisfy:

1. $P(x_1, x_2, \ldots, x_n) \geq 0$

2. $\displaystyle\sum_{x_1, \ldots, x_n} P(x_1, x_2, \ldots, x_n) = 1$

Prior or unconditional probabilities of propositions
e.g., P (*Cavity* = *true*) = 0.1 and P (*Weather* = *sunny*) = 0.72
correspond to belief prior to arrival of any (new) evidence

Joint probability distributions

Joint distributions are typically written in table form:

T	W	P(T,W)
Warm	snow	0.1
Warm	hail	0.3
Cold	snow	0.5
Cold	hail	0.1

Marginalization

Given P(T,W), calculate P(T) or P(W)...

T	W	P(T,W)
Warm	snow	0.1
Warm	hail	0.3
Cold	snow	0.4
Cold	hail	0.2

$$P(T) = \sum_{w \in W} P(T, w)$$

T	P(T)
Warm	0.4
Cold	0.6

$$P(W) = \sum_{t \in T} P(t, W)$$

W	P(W)
snow	0.5
hail	0.5

Marginalization

Given P(T,W), calculate P(T) or P(W)...

T	W	P(T,W)
Warm	snow	0.3
Warm	hail	0.2
Cold	snow	0.2
Cold	hail	0.3

$$P(T) = \sum_{w \in W} P(T, w)$$

T	P(T)
Warm	?
Cold	?

$$P(W) = \sum_{t \in T} P(t, W)$$

W	P(W)
snow	?
hail	?

Conditional Probabilities

Conditional or posterior probabilities

- e.g., $P(cavity|toothache) = 0.8$
- i.e., given that toothache is all I know

If we know more, e.g., cavity is also given, then we have $P(cavity|toothache, cavity) = 1$

- Note: the less specific belief remains valid after more evidence arrives, but is not

always useful

New evidence may be irrelevant, allowing simplification

- e.g., $P(cavity|toothache, redsoxwin)=P(cavity|toothache)=0.8$

This kind of inference, sanctioned by domain knowledge, is crucial

Conditional Probabilities

Conditional or posterior probabilities

- e.g., *P*(*cavity*|*toothache*) = 0.8
- i.e., given that toothache is all I know

Often written as a conditional probability table:

cavity	P(cavity|toothache)
true	0.8
false	0.2

If we know more, e.g., cavity is also give

- Note: the less specific belief rema

always useful

New evidence may be irrelevant, allowing simplification

- e.g., *P*(*cavity*|*toothache*, *redsoxwin*)=*P*(*cavity*|*toothache*)=0.8

This kind of inference, sanctioned by domain knowledge, is crucial

Conditional Probabilities

Conditional probability: $P(A \mid B) = \dfrac{P(A,B)}{P(B)}$ (if P(B)>0)

Example: Medical diagnosis

Product rule: P(A,B) = P(A ∧ B) = P(A|B)P(B)

Marginalization with conditional probabilities:

$$P(A) = \sum_{b \in B} P(A \mid B = b) P(B = b)$$

This formula/rule is called the *law of of total probability*

Chain rule is derived by successive application of product rule:

$P(X_1,\dots,X_n) = P(X_1,\dots,X_{n-1})\ P(X_n|X_1,\dots,X_{n-1})$

$= P(X_1,\dots,X_{n-2})\ P(X_{n-1}|X_1,\dots,X_{n-2})\ P(X_n|X_1,\dots,X_{n-1}) = \dots$

$= \Pi^n_{i=1}\ P(X_i|X_1,\dots,X_{i-1})$

Conditional Probabilities

P(snow|warm) = Probability that it will snow *given* that it is warm

T	W	P(T,W)
Warm	snow	0.3
Warm	hail	0.2
Cold	snow	0.2
Cold	hail	0.3

Conditional distribution

Given P(T,W), calculate P(T|w) or P(W|t)...

T	W	P(T,W)
Warm	snow	0.3
Warm	hail	0.2
Cold	snow	0.2
Cold	hail	0.3

W	P(W|T=warm)
snow	?
hail	?

$$P(W|t) = \frac{P(W,t)}{P(t)}$$

Conditional distribution

Given P(T,W), calculate P(T|w) or P(W|t)...

T	W	P(T,W)
Warm	snow	0.3
Warm	hail	0.2
Cold	snow	0.2
Cold	hail	0.3

W	P(W|T=warm)
snow	?
hail	?

$$\boxed{P(W|t) = \frac{P(W,t)}{P(t)}}$$

Where did this formula come from?

Conditional distribution

Given P(T,W), calculate P(T|w) or P(W|t)...

T	W	P(T,W)
Warm	snow	0.3
Warm	hail	0.2
Cold	snow	0.2
Cold	hail	0.3

W	P(W|T=warm)
snow	?
hail	?

$$P(W|t) = \frac{P(W,t)}{P(t)}$$

$$P(snow|warm) = \frac{P(warm, snow)}{P(warm)} = \frac{P(warm, snow)}{P(warm, hail) + P(warm, snow)}$$

Conditional distribution

Given P(T,W), calculate P(T|w) or P(W|t)...

T	W	P(T,W)
Warm	snow	0.3
Warm	hail	0.2
Cold	snow	0.2
Cold	hail	0.3

W	P(W|T=warm)
snow	0.6
hail	?

$$P(W|t) = \frac{P(W,t)}{P(t)}$$

$$P(snow|warm) = \frac{P(warm, snow)}{P(warm)} = \frac{P(warm, snow)}{P(warm, hail) + P(warm, snow)}$$

$$= \frac{0.3}{0.2 + 0.3}$$

Conditional distribution

Given P(T,W), calculate P(T|w) or P(W|t)...

T	W	P(T,W)
Warm	snow	0.3
Warm	hail	0.2
Cold	snow	0.2
Cold	hail	0.3

W	P(W|T=warm)
snow	0.6
hail	?

How do we solve for this?

$$P(W|t) = \frac{P(W,t)}{P(t)}$$

$$P(snow|warm) = \frac{P(warm, snow)}{P(warm)} = \frac{P(warm, snow)}{P(warm, hail) + P(warm, snow)}$$

$$= \frac{0.3}{0.2 + 0.3}$$

Conditional distribution

Given P(T,W), calculate P(T|w) or P(W|t)...

T	W	P(T,W)
Warm	snow	0.3
Warm	hail	0.2
Cold	snow	0.2
Cold	hail	0.3

W	P(W|T=warm)
snow	0.6
hail	0.4

$$P(W|t) = \frac{P(W,t)}{P(t)}$$

$$P(snow|warm) = \frac{P(warm, snow)}{P(warm)} = \frac{P(warm, snow)}{P(warm, hail) + P(warm, snow)}$$

$$= \frac{0.3}{0.2 + 0.3}$$

Conditional distribution

Given P(T,W), calculate P(T|w) or P(W|t)...

T	W	P(T,W)
Warm	snow	0.3
Warm	hail	0.2
Cold	snow	0.2
Cold	hail	0.3

W	P(W|T=warm)
snow	0.6
hail	0.4

$$P(W|t) = \frac{P(W,t)}{P(t)}$$

W	P(W|T=cold)
snow	?
hail	?

Conditional distribution

Given P(T,W), calculate P(T|w) or P(W|t)...

T	W	P(T,W)
Warm	snow	0.3
Warm	hail	0.2
Cold	snow	0.2
Cold	hail	0.3

$$P(W|t) = \frac{P(W,t)}{P(t)}$$

W	P(W|T=warm)
snow	0.6
hail	0.4

W	P(W|T=cold)
snow	0.4
hail	0.6

Normalization

T	W	P(T,W)
Warm	snow	0.3
Warm	hail	0.2
Cold	snow	0.2
Cold	hail	0.3

W	P(W|T=warm)
snow	0.6
hail	0.4

$$P(W|t) = \frac{P(W,t)}{\boxed{P(t)}}$$

Can we avoid explicitly computing this denominator?

$$P(snow|warm) = \frac{P(warm, snow)}{\boxed{P(warm, hail) + P(warm, snow)}}$$

Any ideas?

Normalization

T	W	P(T,W)
Warm	snow	0.3
Warm	hail	0.2
Cold	snow	0.2
Cold	hail	0.3

$$P(W|t) = \frac{P(W,t)}{P(t)}$$

W	P(W|T=warm)
snow	0.6
hail	0.4

Two steps:

1. Copy entries

W	P(W,T=warm)
snow	0.3
hail	0.2

2. Scale them up so that entries sum to 1

W	P(W|T=warm)
snow	0.6
hail	0.4

Normalization

T	W	P(T,W)
Warm	snow	0.3
Warm	hail	0.4
Cold	snow	0.2
Cold	hail	0.1

<u>Two steps:</u>

1. Copy entries

T	P(T,W=hail)
warm	?
cold	?

2. Scale them up so that entries sum to 1

T	P(T|W=hail)
warm	?
cold	?

Normalization

T	W	P(T,W)
Warm	snow	0.3
Warm	hail	0.4
Cold	snow	0.2
Cold	hail	0.1

<u>Two steps:</u>

1. Copy entries

T	P(T,W=hail)
warm	0.4
cold	0.1

2. Scale them up so that entries sum to 1

T	P(T|W=hail)
warm	?
cold	?

Normalization

T	W	P(T,W)
Warm	snow	0.3
Warm	hail	0.4
Cold	snow	0.2
Cold	hail	0.1

Two steps:

1. Copy entries

T	P(T,W=hail)
warm	0.4
cold	0.1

2. Scale them up so that entries sum to 1

T	P(T|W=hail)
warm	0.8
cold	0.2

$$P(W|t) = \frac{P(W,t)}{P(t)}$$

The only purpose of this denominator is to make the distribution sum to one.
– we achieve the same thing by scaling.

Bayes Rule

$$P(a|b) = \frac{P(b|a)P(a)}{P(b)}$$

<u>Thomas Bayes (1701 – 1761)</u>:
– English statistician, philosopher and Presbyterian minister
– formulated a specific case of the formula above
– his work later published/generalized by Richard Price

Bayes Rule

$$P(a|b) = \frac{P(b|a)P(a)}{P(b)}$$

It's easy to derive from the product rule:

$$P(a,b) = P(b|a)P(a) = \underbrace{P(a|b)}_{\text{Solve for this}}P(b)$$

Using Bayes Rule

$$P(a|b) = \frac{P(b|a)P(a)}{P(b)}$$

$$P(cause|effect) = \frac{P(effect|cause)P(cause)}{P(effect)}$$

Using Bayes Rule

$$P(a|b) = \frac{P(b|a)P(a)}{P(b)}$$

$$P(cause|effect) = \frac{\boxed{P(effect|cause)}P(cause)}{P(effect)}$$

But harder to estimate this

It's often easier to estimate this

Bayes Rule Example

$$P(cause|effect) = \frac{P(effect|cause)P(cause)}{P(effect)}$$

Suppose you have a stiff neck...

Suppose there is a 70% chance of meningitis if you have a stiff neck:

stiff neck meningitis

$$P(s|m) = 0.7$$

What are the chances that you have meningitis?

Bayes Rule Example

$$P(cause|effect) = \frac{P(effect|cause)P(cause)}{P(effect)}$$

Suppose you have a stiff neck...

Suppose there is a 70% chance of meningitis if you have a stiff neck:

$$P(\overset{\text{stiff neck}}{s}|\overset{\text{meningitis}}{m}) = 0.7$$

What are the chances that you have meningitis?

We need a little more information...

Bayes Rule Example

$$P(cause|effect) = \frac{P(effect|cause)P(cause)}{P(effect)}$$

$$P(s|m) = 0.7$$

$$P(s) = 0.01$$

Prior probability of stiff neck

$$P(m) = \frac{1}{50000}$$

Prior probability of meningitis

$$P(m|s) = \frac{P(s|m)P(m)}{P(s)} = \frac{0.7 \times \frac{1}{50000}}{0.01} = 0.0014$$

Bayes Rule Example

Given:

W	P(W)
snow	0.8
hail	0.2

T	W	P(T|W)
Warm	snow	0.3
Warm	hail	0.4
Cold	snow	0.7
Cold	hail	0.6

Calculate P(W|warm):

$$P(W|warm) = \frac{P(warm|W)P(W)}{P(warm)}$$

Bayes Rule Example

Given:

W	P(W)
snow	0.8
hail	0.2

T	W	P(T|W)
Warm	snow	0.3
Warm	hail	0.4
Cold	snow	0.7
Cold	hail	0.6

Calculate P(W|warm):

$$P(W|warm) = \frac{P(warm|W)P(W)}{P(warm)}$$

$$P(hail|warm) = \frac{0.4 \times 0.2}{P(warm)} = \frac{0.08}{P(warm)}$$

$$P(snow|warm) = \frac{0.3 \times 0.8}{P(warm)} = \frac{0.24}{P(warm)}$$

normalize

=0.25

=0.75

Independence

If two variables are independent, then: $P(a, b) = P(a)P(b)$

or

$$P(a) = P(a|b)$$

or

$$P(b) = P(b|a)$$

Independence

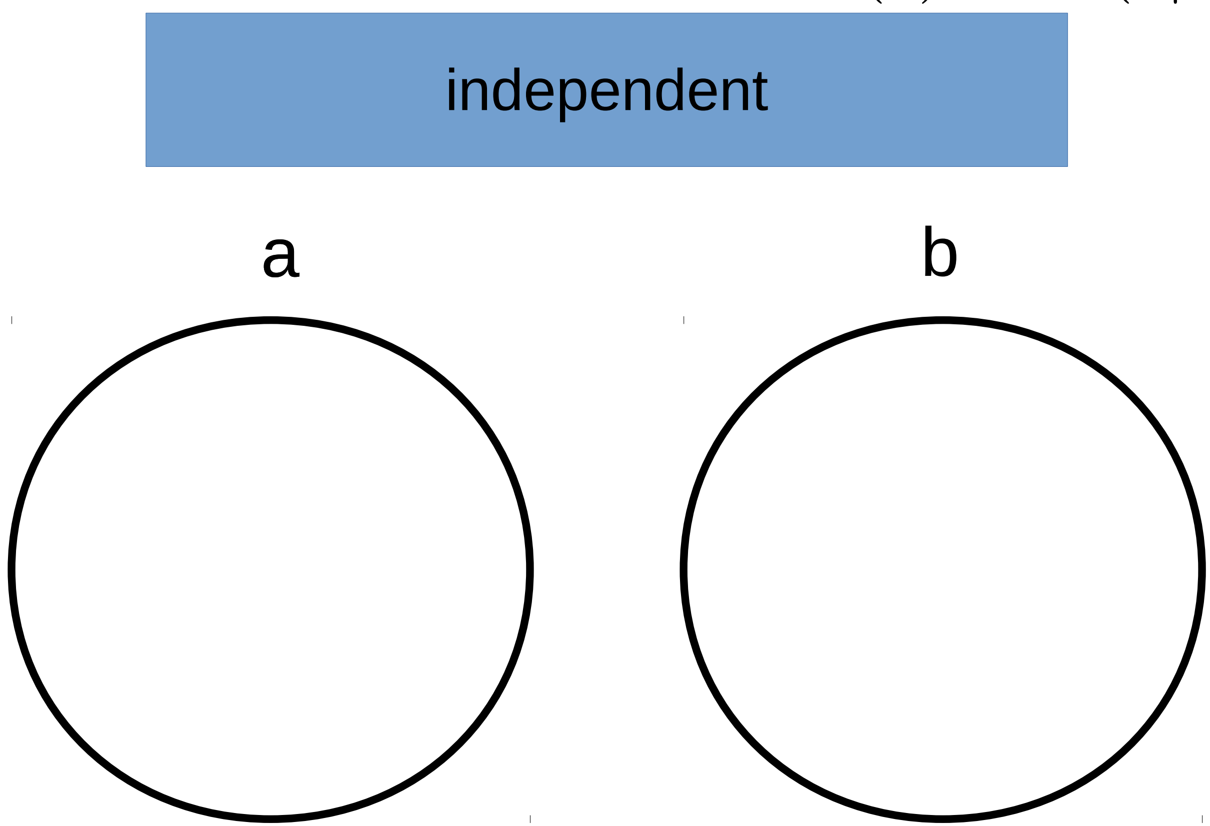
If two variables are independent, then: $P(a,b) = P(a)P(b)$

or

$$P(a) = P(a|b)$$

or

$$P(b) = P(b|a)$$



Independence

If two variables are independent, then: $$P(a,b) = P(a)P(b)$$

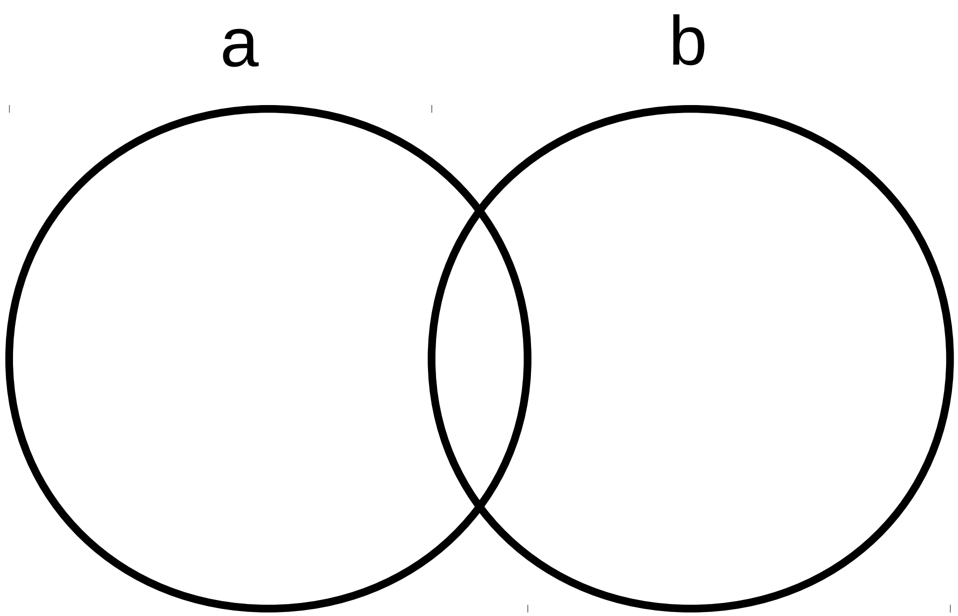
or

$$P(a) = P(a|b)$$

or

$$P(b) = P(b|a)$$

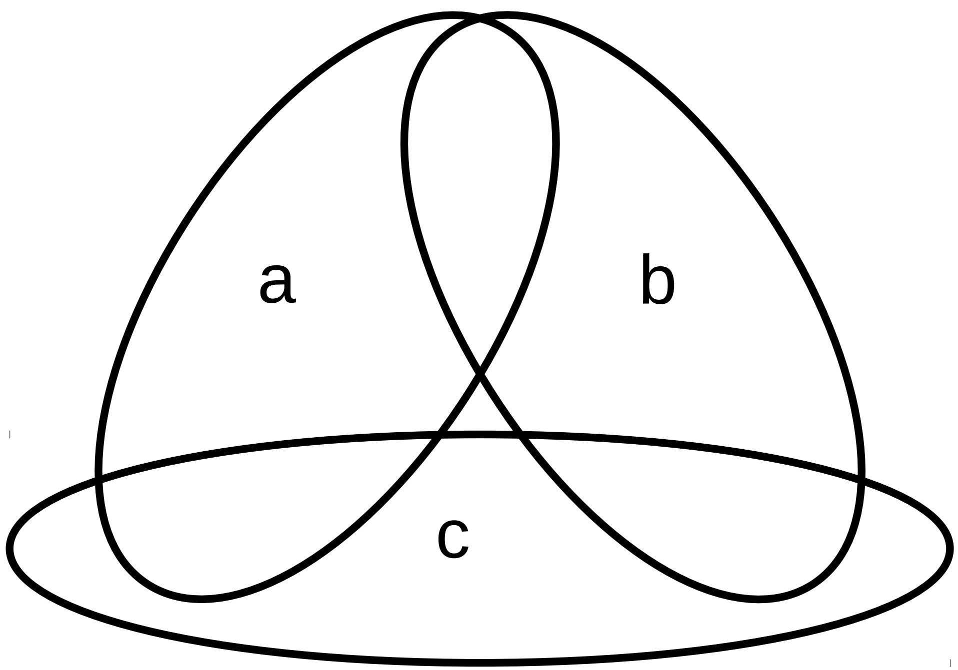
Not independent



Conditional Independence

If two variables a,b are *conditionally* independent given c, then:

$$P(a,b|c) = P(a|c)P(b|c)$$



Without conditioning on c, a and b are not independent!!!