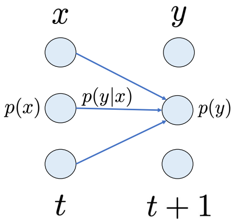
Marginal, conditional and joint distributions



Marginal: $p_t(x) = P(X_t = x)$, $p_{t+1}(y) = P(X_{t+1} = y)$.
Conditional: Forward $p(y|x) = P(X_{t+1} = y|X_t = x)$.
x: cause, y: effect. $p(y|x)$: cause $\rightarrow$ effect, given or learned.
Joint: $p(x, y) = P(X_t = x, X_{t+1} = y)$.
Chain rule: $p(x, y) = p_t(x)p(y|x)$.
Rule of total probability:
$p_{t+1}(y) = \sum_x p(x, y) = \sum_x p_t(x)p(y|x)$.
Add up probabilities of alternative chains of events.

Disease and symptom

100A

Xiaowu Dai

Basics
Population
Region
Coin
Markov
Reasoning

Example 6: Rare disease example

1% of population has a rare disease.

A random person goes through a test.

If the person has disease, 90% chance test positive.

If the person does not have disease, 90% chance test negative.

If tested positive, what is the chance he or she has disease?

$P(D) = 1\%$.

Forward: $P(+|D) = 90\%$, $P(-|N) = 90\%$.

Backward: $P(D|+) = ?$

Example 6: Rare disease example

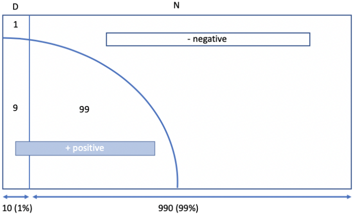
$P(D) = 1\%$.

Forward: from cause to effect.

$P(+|D) = 90\%$, $P(-|N) = 90\%$.

Backward: from effect to cause.

$P(D|+) = ?$

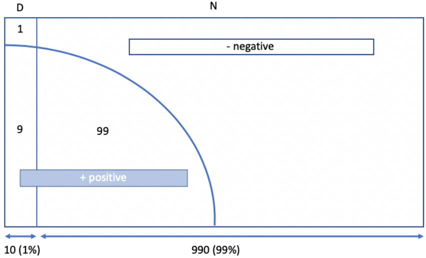


$P(D|+) = \frac{9}{9+99} = \frac{1}{12}$.

$P(\text{alarm} \mid \text{fire})$ vs $P(\text{fire} \mid \text{alarm})$.

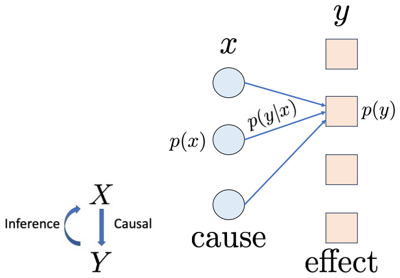
Example 6: Rare disease example

$P(D) = 1\%$.
$P(+|D) = 90\%$, $P(-|N) = 90\%$.



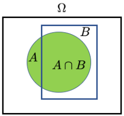
$P(D \cap +) = P(D)P(+|D) = 1\% \times 90\%$.
$P(N \cap +) = P(N)P(+|N) = 99\% \times 10\%$.
$P(+) = P(D \cap +) + P(N \cap +) = 1\% \times 90\% + 99\% \times 10\%$.
$P(D|+) = \frac{P(D \cap +)}{P(+)} = \frac{9}{9+99} = \frac{1}{12}$.

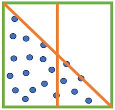
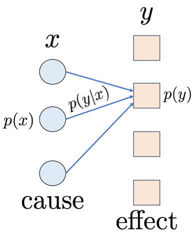
Random variables, probability mass functions



Marginal: prior $p(x) = P(X = x)$, marginal $p(y) = P(Y = y)$.
Conditional: forward generation $p(y|x) = P(Y = y|X = x)$
backward inference $p(x|y) = P(X = x|Y = y)$.
Chain rule: joint $p(x, y) = p(x)p(y|x)$.
Rule of total probability: marginal
$p(y) = \sum_x p(x, y) = \sum_x p(x)p(y|x)$.

Bayes rule



$$P(A|B) = \frac{P(A \cap B)}{P(B)}.$$

Bayes rule: backward inference, back tracing, posterior

$$p(x|y) = P(X = x|Y = y) = \frac{P(X = x, Y = y)}{P(Y = y)}$$

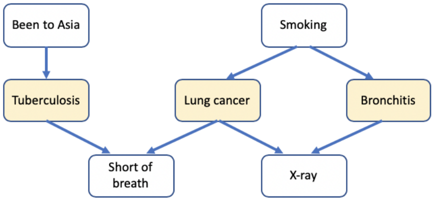
$$= \frac{p(x, y)}{p(y)} = \frac{p(x)p(y|x)}{\sum_{x'} p(x')p(y|x')}.$$

Conditional:

(1) **Forward:** cause $\rightarrow$ effect, physical, given. fire $\rightarrow$ alarm.

(2) **Backward:** effect $\rightarrow$ cause, mental, inferred. alarm $\rightarrow$ fire.

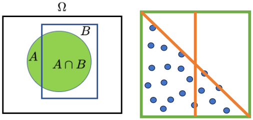
Bayes network, directed acyclic graph, graphic model



Conditional independence:

(1) Sibling nodes are independent given parent node.

(2) Child node is independent of grandparents given parent.



$$P(A|B) = \frac{P(A \cap B)}{P(B)}.$$

$$P(A \cap B) = P(B)P(A|B).$$

Independence

$$P(A|B) = P(A).$$

$$P(A \cap B) = P(A)P(B).$$

A and B have nothing to do with each other.

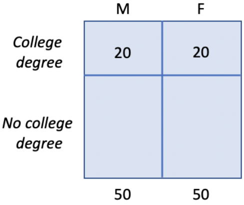
Independence

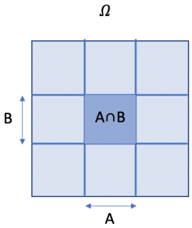
Definition 1:

$$P(A|B) = P(A).$$
$$p(y|x) = p(y).$$

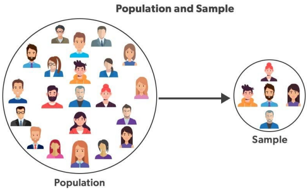
Definition 2:

$$P(A \cap B) = P(A)P(B).$$
$$p(x, y) = p(x)p(y).$$





Population of sequences



Sample a person from population Ω_1 of N people uniformly.
Repeat n times independently.
$\Omega_n = \{$ all N^n possible sequences $\}$.
equally likely outcomes in Ω_1 + independent repetitions
= equally likely sequences in Ω_n.
Let $\omega = (a_1, a_2, ..., a_n) \in \Omega_n$, each $a_i \in \Omega_1$.
$P(\omega) = P(a_1)P(a_2)...P(a_n) = \frac{1}{N} \times \frac{1}{N} \times ... \times \frac{1}{N} = \frac{1}{N^n}$.
Coin flipping: $\Omega_1 = \{$ head, tail $\}$.
Die rolling: $\Omega_1 = \{1, 2, ..., 6\}$.
Uniform random number $\Omega_1 = [0, 1]$.

Conditional independence

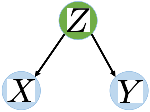
Markov chain: $C \to B \to A$, $Z \to X \to Y$.

$$P(A|B,C) = P(A|B).$$
$$p(y|x,z) = p(y|x).$$

Future is independent of the past given present.
Immediate cause (parent), remote cause (grandparent).
Meta rule: Insert same condition in a definition or equation.

Conditional independence

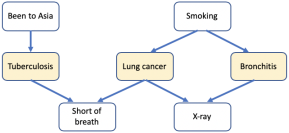
Shared cause: $C \leftarrow B \rightarrow A$.



$$P(A \cap C|B) = P(A|B)P(C|B).$$
$$p(x, y|z) = p(x|z)p(y|z).$$

Children given parent.
Meta rule: Insert same condition in a definition or equation.

Bayes net



a: Been to Asia; s: Smoking; t: Tuberculosis; l: Lung cancer; b: Bronchitis; d: Short of breath (Dyspnea); x: X-ray.

$$p(a,s,t,l,b,d,x) = p(a)p(s)p(t|a)p(l|s)p(b|s)p(d|t,l)p(x|b,l),$$

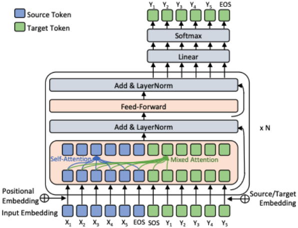
$$p(l|a,s,d,x) = \frac{p(l,a,s,d,x)}{p(a,s,d,x)},$$

$$p(l,a,s,d,x) = \sum_{t,b} p(a,s,t,l,b,d,x),$$

$$p(a,s,d,x) = \sum_{l} p(l,a,s,d,x).$$

Efficient calculation: message passing / belief propagation.

Generative Pre-trained Transformer (GPT)

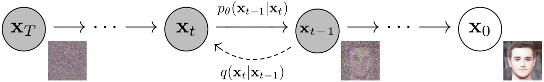


$x = (x_1, ..., x_{T_x})$ (e.g., "Can you write a poem?")
$y = (y_1, ..., y_{T_y})$ (e.g., "Certainly. Below is the poem...")
$p(y|x) = \prod_{t=1}^{T_y} p(y_t|y_{<t}, x)$.
Learn from training data $(x^{(i)}, y^{(i)}, i = 1, ..., n)$ by maximizing

$$\frac{1}{n}\sum_{i=1}^{n} \log p_\theta(y^{(i)}|x^{(i)}) = \frac{1}{n}\sum_{i=1}^{n}\sum_{t} \log p_\theta(y_t^{(i)}|y_{<t}^{(i)}, x^{(i)}).$$

memorize and generalize (interpolation).

Denoising Diffusion Probability Model



x_0: clean image.

$x_t = x_{t-1} + e_t$, e_t: small noise

Forward noising $q(x_t|x_{t-1})$, $t = 1, ..., T$. x_T: big noise.

Backward denoising $p(x_{t-1}|x_t)$.

Learn from training data $(x_0^{(i)}, i = 1, ..., n)$ by maximizing

$$\frac{1}{n}\sum_{i=1}^{n}\sum_{t=T}^{1}\log p_\theta(x_{t-1}^{(i)}|x_t^{(i)}).$$

memorize and generalize (interpolation).

Take home message

As long as you can count

Count the population (of equally likely outcomes)

Count the repetitions (sequence of outcomes, fluctuation)

Population of sequences of repetitions (equally likely sequences)

Population of trajectories (random walk)

Two things

(1) Intuition, visualization and motivation

(2) Precise notation and formula

Accomplished

Most of the important concepts via intuitive examples

Next

Systematic and more in-depth treatments

Random variables and probability functions, expectation

Continuous random variables, continuous time processes