

today:

- motivation
- prob. review

why graphical models?

PoS tagging observation  $(x_1, x_2, x_3, \dots, x_T) \triangleq x_{1:T}$

$$x_t \in \{1, \dots, k\}$$

want to model  $p(x_{1:T})$  issue: exponential size of state space  
Length input

$\Rightarrow k^T - 1$  parameters needed to fully describe distribution

trick: make a factorization assumption about  $p$

$$p(x_1, \dots, x_T) = f_1(x_1) f_2(x_2|x_1) \dots f_T(x_T|x_{T-1})$$

factor → 2 variables  $\Rightarrow \approx k^2$  parameters to specify  
T-factors  $\Rightarrow$   $T \cdot k^2$  parameters  $\ll k^T$

clique  $\Rightarrow$   
in a graph (graphical model)  
theme: representation

computation?

say want to compute  $p(x_1)$  "marginal"

$$p(x_1) = \sum_{x_2, x_3, \dots, x_T} p(x_{1:T})$$

distributivity  
 $a \cdot (b+c) = a \cdot b + a \cdot c$

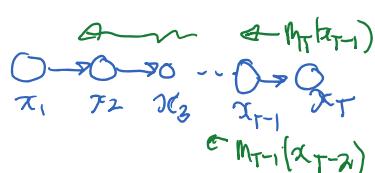
exponential # of terms

$$\sum_{x_2:T} \sum_{x_2 \in \{1, \dots, k\}} \sum_{x_3 \in \{1, \dots, k\}} \dots$$

$$= f_1(x_1) \left( \sum_{x_2} f_2(x_2|x_1) \left( \sum_{x_3} f_3(x_3|x_2) \left( \dots \left( \sum_{x_T} f_T(x_T|x_{T-1}) \dots \right) \right) \right) \right)$$

$$\leq S_{T-1}(x_{T-1}|x_{T-2}) \cdot m_{T-1}(x_{T-1})$$

$m_{T-1}(x_{T-1}) \in O(k^2)$   
bcapit



"message passing alg." to compute efficiently marginal  $p(x_1)$

$\rightarrow O(T \cdot k^2)$  time

### key themes:

I) representation: how to represent structured prob. dist?

(probability)

- graph  $\rightsquigarrow$  factorization
- parametrization  $\rightsquigarrow$  full table

"exponential family"

II) estimation: given data, how to learn/estimate the parameters of dist.

(statistics)

- $\rightsquigarrow$  learning e.g.:
- maximum likelihood
  - max. entropy
  - moment matching

III) "probabilistic"

(CS)

inference: answer questions about data

e.g. compute  $p(y|x)$

conditional  
observation  
query

or  $p(x)$   
marginal

$\rightsquigarrow$  Computation e.g.:

- message passing
- approximate inference
  - sampling
  - variational methods

### next: probability review

why?  $\rightarrow$  principled framework to model uncertainty

#### Sources of uncertainty

1) intrinsic uncertainty  $\rightarrow$  quantum mechanics

2) partial information / observation :

- card games
- rolling a die
- $\rightarrow$  don't know the initial conditions "exactly enough"

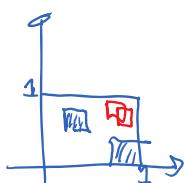
3) incomplete modeling of a complex phenomenon

(computational issues are also important)

example: "most birds can fly"  
 $\rightarrow$  simple rule can be advantageous but then yields uncertainty

• "AI"

Probabilities  $\rightsquigarrow$  like areas



16h02

Notebook:  $X_1 \ X_2 \ X_3 \quad X \ Y \ Z$

{ random variables (usually real valued) }

- - - - -

$x_1, x_2, x_3 \dots x y z$   
their "realizations"

$X$  a random variable  $\rightarrow$  represents an uncertainty quantity

$X=x$  represents the "event" that  $X$  takes the value  $x$

$\Omega \rightarrow$  the sample space of "elementary events"  
possible values of my R.V.s

e.g.  $X$   
"result of a die"  
roll of a die

$$\Omega = \{1, 2, 3, 4, 5, 6\}$$

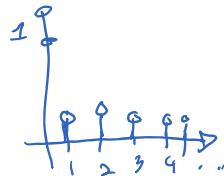
two types of R.V.  $\leftarrow$  discrete where  $\Omega$  is countable  
continuous "  $\Omega$  " uncountable

I) [assume  $\Omega$  is countable  $\rightarrow$  discrete case]

(discrete R.V. is characterized by a probability mass function (pmf))

"lower case"  $\rightarrow p(x)$  for  $x \in \Omega$

$$\text{pmf } p \text{ is s.b. } \begin{cases} p(x) \geq 0 & \forall x \in \Omega \\ \sum_{x \in \Omega} p(x) = 1 \end{cases}$$



probability distribution  $P$

is a mapping  $P: E \rightarrow [0, 1]$

$E = 2^{\Omega} \triangleq$  set of all subsets of  $\Omega$

that satisfies the Kolmogorov axioms  $P(E) \geq 0 \quad \forall E \in E$

$$P(\Omega) = 1, \text{ countable}$$

$$P(\bigcup_{i=1}^{\infty} E_i) = \sum_{i=1}^{\infty} P(E_i) \quad \text{when } E_i \text{ are mutually disjoint}$$

("σ-field" in measure theory  
needed when  $\Omega$  is uncountable)

notation:  $P\{X=x\} = p(x)$

$$P\{X=x \text{ or } X=x'\} = P\{X=x\} + P\{X=x'\} = p(x) + p(x') \quad \text{if } x \neq x'$$

$$\{w : X(w) = x\}$$

$$E = \{x, x'\}$$

$$\text{for discrete R.V. } P(E) = \sum_{x \in E} p(x)$$

continuous R.V.

continuous R.V. is characterized by a probability density function  $p$  (pdf)

$$p: \Omega \rightarrow \mathbb{R}$$

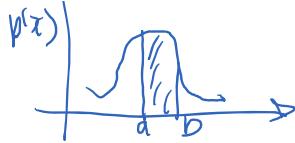
$$p(x) \geq 0 \quad \forall x \in \Omega$$

$$p \text{ is integrable and } \int p(x) dx = 1$$



$$p(x) \geq 0 \quad \forall x \in \mathbb{L}$$

$p$  is integrable and  $\int_{\mathbb{L}} p(x) dx = 1$



prob. of events

$$\mathbb{L} = \mathbb{R} : \quad P([a, b]) = \int_a^b p(x) dx$$

$\Sigma \rightarrow$  Borel σ-fields

$$p(x) = \lim_{n \rightarrow \infty} \frac{P(B_n)}{\text{vol}(B_n)}$$

recap: discrete R.V.  $X$ ; pmf  $p(x) \Leftrightarrow P\{X=x\} = p(x)$  where  $B_n$  are balls centered around  $x$  with decreasing radius

cts. R.V.  $X$ ; pdf  $p(x) \quad P\{X=x\} = 0$

" $P\{X \in x \pm \frac{\alpha}{2}\} \approx p(x)dx$ "

$$P\{X \in x \pm \frac{\alpha}{2}\} = \int_{x-\frac{\alpha}{2}}^{x+\frac{\alpha}{2}} p(u) du$$

[ sidenote: can change pdf  $p$  on a countable # of pts without changing anything ]

has measure zero according to Lebesgue measure

Joint, marginal etc. (context: multivariate R.V.)

$$Z = (X, Y) \quad \mathbb{L}_Z \triangleq \mathbb{L}_X \times \mathbb{L}_Y$$

(discrete) pmf of  $Z$  "joint pmf" on  $X \times Y$

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

↑ joint distribution

can represent elementary events of  $\mathbb{L}_{(X,Y)}$  as table

	$X=0$	$X=1$
$Y=0$	○	○
$Y=1$	○	○

$p(1,1)$

say  $X$  is whether die result is even  
 $Y$  " " " " " odd

cts case: if  $X \times Y$  are cts "and"  $P\{\text{box region}\} = \int_{\text{box}} p(x, y) dx dy$

marginal distribution: (in a context of a joint dist.)

↳ dist. of a component of a random vector

$$P\{X=x\} = \sum_y P\{X=x, Y=y\} \quad \text{"sum rule"}$$

$$\sim p(x,y) \xrightarrow{\text{"marginalizing out $x$"}}$$