



- approximate inference
  - sampling
  - variational methods

today: probability review

why? → principled framework to model uncertainty

sources of uncertainty

1) intrinsic uncertainty → quantum mechanics

2) partial information / observation :

- card games

- rolling a dice

→ don't know the initial conditions

3) incomplete modeling of a complex phenomenon

example: "most birds can fly"

→ simple rule can be advantageous but then yields uncertainty

(computational issues are also important)

- "AI"

Notation :

$X_1, X_2, X_3$

$X, Y, Z$

↑  
random variables

(usually real-valued)

$x_1, x_2, x_3$

$x, y, z$

↑ their "realizations"

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

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

$\Omega \rightarrow$  the sample space of "elementary events"

possible values for my R.V.

e.g.  $X$  is result of a die throw

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

two types of R.V.   
  $\left\{ \begin{array}{l} \text{discrete} \text{ where } \Omega \text{ is countable} \\ \text{continuous} \text{ (} \Omega \text{ is uncountable)} \end{array} \right.$

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

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

"point case"

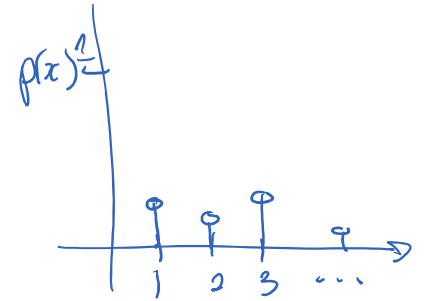
$\rightarrow p(x)$  for  $x \in \Omega$

$$p(x) = \frac{1}{4}$$

"small case"

$$p(x) \text{ for } x \in \Omega$$

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



probability distribution

big  $\mathcal{P}$

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

that satisfy the Kolmogorov axioms

$$P(E) \geq 0 \quad \forall E \in \mathcal{E}$$

$$P(\Omega) = 1$$

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

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

set of "events"

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

notation:  $P\{X=x\}$

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

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

$E = \{x, y\}$       pmf

14h33

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

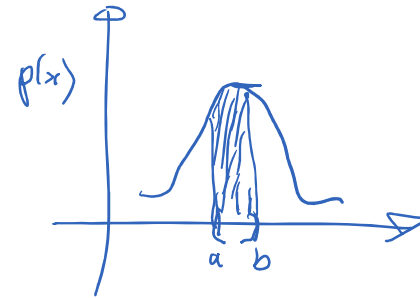
Continuous R.V. is characterized by a probability density function  $p(x)$  (ndf)

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

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

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

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



prob. of events:

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



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

where  $B_n$  are balls enclosing  $x$  and decreasing in radius

Recap:

discrete R.V.  $X$ ; pmf  $p(x) \Leftrightarrow P\{X=x\} = p(x)$

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

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

$$P\{X \in x \pm \frac{\Delta}{2}\} = \int_{x-\frac{\Delta}{2}}^{x+\frac{\Delta}{2}} p(x) dx$$

[side note: can change pdf  $p$  on countable # of pts. *has measure zero according to Lebesgue*

[side note: can change pdf  $p$  on countable # of pts. <sup>according to Lebesgue measure</sup> without changing "anything" ]

joint, marginal, etc. (multivariate R.V.)

$$Z = (X, Y) \quad \Omega_Z \triangleq \Omega_X \times \Omega_Y$$

(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  $\Omega_{(X,Y)}$  as table

|       | $X=0$         | $X=1$         |
|-------|---------------|---------------|
| $Y=0$ | 0             | $\frac{1}{2}$ |
| $Y=1$ | $\frac{1}{2}$ | 0             |

↙  $p(1,1)$

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

cts. case: if  $X$  &  $Y$  are cts <sup>"and"</sup>

$$P\{\text{box region}\} = \int_{\text{box}} p(x, y) dx dy$$

↙ joint pdf

marginal distribution (in the context of a joint)

↳ dist. of a component of a random vector

$$P\{X=x\} = \sum_{y \in \Omega_Y} \frac{P\{X=x, Y=y\}}{p(x,y)} \quad \text{"sum rule"}$$

"marginalizing out Y"

Independence:

•  $X$  is independent of  $Y \iff p(x,y) = p(x)p(y) \quad \forall (x,y) \in \Omega_X \times \Omega_Y$

notation:  $X \perp\!\!\!\perp Y$

• R.V.  $X_1, \dots, X_n$  are "mutually independent"  $\iff p(x_{1:n}) = \prod_{i=1}^n p(x_i) \quad \forall x_{1:n} \in \prod_{i=1}^n \Omega_{X_i}$

conditioning:

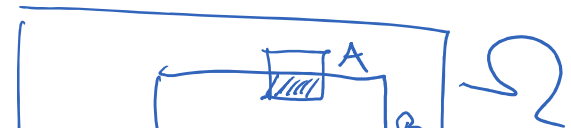
• for events  $A \& B$ , suppose  $P(B) \neq 0$

then define  $P(A|B) \triangleq \frac{P(A \cap B)}{P(B)}$

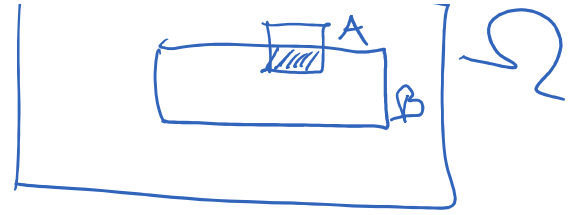
↑  
normalize

illustration:

$$P(B) = \sum_{A \in \text{partition of } \Omega} P(A \cap B)$$



normalizen



- for discrete R.V. define "conditional pmf"

$$p(x|y) \triangleq P(X=x|Y=y) \triangleq \frac{P(\{X=x, Y=y\}) \text{ \{joint pmf\}}}{P(\{Y=y\}) \text{ \{marginal pmf\}}} = \frac{p(x,y)}{p(y)}$$

$p(x|y)$  or  $p(x,y)$

(normalization cond.:  $\sum_x p(x,y) = p(y)$ )

- for cts. R.V.

"conditional pdf"  $p(x|y) \triangleq \frac{p(x,y)}{p(y)}$  density

subtle point:  $p(x|y)$  is undefined when  $p(y)=0$

[ independence: if  $X \perp Y$   $p(x|y) = \frac{p(x,y)}{p(y)} \stackrel{\text{indep.}}{=} \frac{p(x)p(y)}{p(y)} = p(x)$  ]



example of conditioning : " $P(\text{having cancer} \mid \text{tumor measurement} = a)$ "