Lecture 2 - prob. review

Monday, September 9, 2024 12:32 PM

today: • themes
• prob. review

Key themes:

I) representation : how to represent structured prob. dist.?
(probability)
• graph ⇝ factorization
• parameterization ⟶ full table
vs.
"exponential family"

II) estimation (statistics) : given data, how to learn/estimate the parameters of dist.
⇝ learning e.g. maximum likelihood
max. entropy
moment matching

III) "probabilistic" inference (CS) : answer questions about data
e.g. compute $p(y|x)$ (conditional) or $p(x)$ (marginal)
y: query, x: observation
→ computation e.g. • message passing
• approximate inference
- sampling
- variational methods

probability review

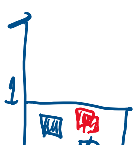
why? ⇝ principled framework to model uncertainty

sources of uncertainty

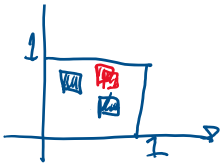
1) intrinsic uncertainty → quantum mechanics
2) partial information/observation : • card game
• rolling a die
→ don't know the initial conditions "exactly enough"
3) incomplete modeling of a complex phenomenon
(computational issues are also important)

example: • "most bird can fly"
→ simple rule can be advantageous but then yields uncertainty

probabilities ⇝ like areas



probabilities ⇝ area



Notation:

$X_1 \quad X_2 \quad X_3 \qquad X \quad Y \quad Z$

random variables (usually real-valued)

$x_1 \quad x_2 \quad x_3 \qquad x \quad y \quad z$

their realizations

X is a random variable → represents an uncertain quantity e.g. X is "result of a roll of a die"

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

Ω → the sample space of "elementary events" possible values of my R.V. $\qquad \Omega=\{1,2,3,4,5,6\}$

two types of R.V. — discrete where Ω is countable
— continuous " Ω " uncountable

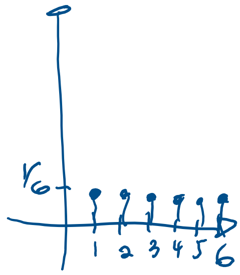
13h28

I) [assume Ω is countable ⇝ discrete case]

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

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

pmf $p: \Omega \to \mathbb{R}_+$ s.t. $\begin{cases} p(x) \geq 0 \ \forall x \in \Omega \\ \sum_{x\in\Omega} p(x) = 1 \end{cases}$



probability distribution "big" P

is a mapping $P: \mathcal{E} \to [0,1]$ $\qquad \mathcal{E} = 2^{\Omega} \triangleq$ set of all subsets of Ω

set of "events"

that satisfies the Kolmogorov axioms: $P(E) \geq 0 \ \forall E \in \mathcal{E}$

$P(\Omega)=1$

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

when E_i's are mutually disjoint

("σ-field" in measure theory)

crucial when Ω is uncountable

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

"formal prob theory"

$P(\{X=x \text{ or } X=x'\}) = P(\{X=x\}) + P(\{X=x'\}) = p(x)+p(x')$ (pmf)

2 here $E = \{x, x'\}$ $\qquad$ $\{x,x\}$ $x \neq x'$

"for meas. prob. theory" $\varnothing$ "$P(\{X=x$ or $X=x'\}) = P(\{X=x\}) + P(\{X=x'\}) = p(x) + p(x')$"

& here $E = \{x, x'\}$ (if $x \neq x'$)

$\{X=x\} \triangleq \{\omega : X(\omega) = x\}$

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

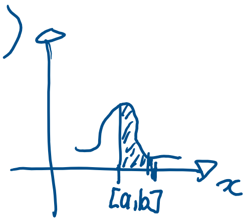
continuous R.V.

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

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

s.t. $p(x) \geq 0 \;\; \forall x \in \Omega$

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



(Ω)

prob. of events

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

$\mathcal{E} \longrightarrow$ Borel σ-field

$p(x) = \lim_{r \to 0^+} \frac{P(B_r)}{vol(B_r)}$ where B_r is a ball of radius r centered at x

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

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

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

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

[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.)
"random vector"

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

(discrete) pmf of Z "joint pmf" on X & Y

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

joint distribution

can represent elementary events of $\Omega_{(X,Y)}$ as a table e.g.

	$X=0$	$X=1$
$Y=0$	0	1/2
$Y=1$	1/2	0

$p(1,1)$

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

cts. case if X & Y are cts ("and")

$$P\{(x,y) \in \text{box region}\} = \int_{\text{box}} \overbrace{p(x,y)}^{\text{joint pdf}}\, dx\, dy$$

marginal distribution (in context of a joint dist.)

↳ dist. of a component of a random vector

$$P\{X=x\} = \sum_{y \in \Omega_Y} \underbrace{P\{X=x, Y=y\}}_{p(x,y)}$$

"sum rule"

"marginalizing out Y"

cts. case joint pdf $p(x,y)$

$$p(x) = \int_y p(x,y)\, dy$$

mixed type: e.g. X is cts, Y is discrete

$p(x,y)$ — cts. → integrate; discrete → sum

$$P\{(x,y) \in \text{set}\} = \sum_y \int_{x \text{ s.t. } (x,y) \in \text{set}} p(x,y)\, dx$$

independence

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

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

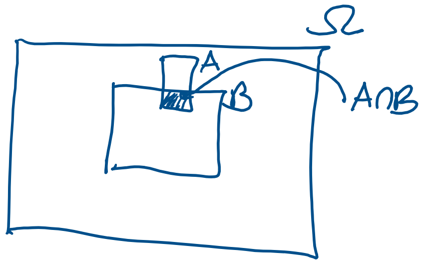
R.V. $X_1, \dots, X_n$ are "mutually independent" $\iff p(x_{1:n}) = \prod_{i=1}^{n} p(x_i)$

$$\forall\, x_{1:n} \in \bigtimes_{i=1}^{n} \Omega_{X_i}$$

conditioning:

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

then $P(A|B) \triangleq \dfrac{P(A \cap B)}{P(B)}$ ← normalization



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

for discrete R.V.

define "conditional pmf"

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

$$p(x|y) \underset{\text{"proportional to"}}{\propto} p(x,y)$$

normalization $\sum_x p(x,y) = p(y)$

for cts. R.V.

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

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

see Borel-Kolmogorov paradox for weird properties of cond. prob.

Borel-Kolmogorov paradox on wikipedia

(*) note: independent $X \perp\!\!\!\perp Y$ $\quad p(x|y) = \frac{p(x,y)}{p(y)} \overset{\text{by } \perp\!\!\!\perp}{=} \frac{p(x)\cancel{p(y)}}{\cancel{p(y)}} = p(x)$