Lecture 3 - scribbles

Tuesday, September 10, 2019 14:29

today:
- finish prob. review
- parametric models

Bayes rule:

→ invert the conditioning

by def.: $p(y|x) = \frac{p(x,y)}{p(x)}$

$\Rightarrow p(x,y) = p(y|x)\,p(x)$ ← "product rule"

$\boxed{\text{Bayes rule}}$

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

chain rule: → successive application of product rule

$$p(x_1,\dots,x_n) = p(x_n|x_{1:n-1})\,\underbrace{p(x_{1:n-1})}$$

$$p(x_n|x_{1:n-1})\,p(x_{n-1}|x_{1:n-2})\,p(x_{1:n-2})$$

...

$$= \prod_{i=1}^{n} \underbrace{p(x_i|x_{1:i-1})}$$

convention: $1{:}0 = \emptyset$, $p(x_1|x_\emptyset) = p(x_1)$

always true? [can choose any $n!$ permutation]

↳ can be simplified when making conditional indep. assumptions (from a directed graphical model)

Conditional independence

X is cond. indep. of Y given Z notation: $X \perp\!\!\!\perp Y \mid Z$

$$\Leftrightarrow p(x,y|z) = p(x|z)\,p(y|z) \quad \forall x \in \Omega_X,\ \forall y \in \Omega_Y,\ \forall z \in \Omega_Z \text{ s.t. } p(z) \neq 0$$

exercise to the reader: prove that if $X \perp\!\!\!\perp Y \mid Z$

then $p(x|y,z) = p(x|z)$

[conditional analog of $X \perp\!\!\!\perp Y \Rightarrow p(x|y) = p(x)$]

example: Z = indicator whether mother carries a genetic disease

X = son 1 has disease $\qquad X \perp\!\!\!\perp Y \mid Z$

example: Z = indicator whether mother carries a genetic disease

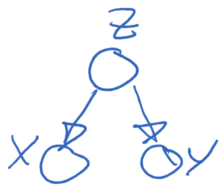
X = son 1 has disease

Y = son 2 " "

$X \perp\!\!\!\perp Y \mid Z$

here, X is ~~not~~ "marginally indep." of Y (i.e. $X \not\perp\!\!\!\perp Y$)

$\exists\, x, y$ s.t. $p(x,y) \neq p(x)\,p(y)$



$$p(x,y,z) = p(x|z)\,p(y|z)\,p(z)$$

example of pairwise indep. R.V.'s that are not mutually independent

X = coin flip as $\{0,1\}$

Y = indep. coin flip

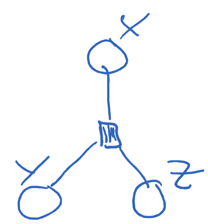
$(X \perp\!\!\!\perp Y)$

define $Z \triangleq X \text{ xor } Y$

$Z \perp\!\!\!\perp X$

$Z \perp\!\!\!\perp Y$

$Z \cancel{\perp\!\!\!\perp} (X, Y)$



(factor graph)

other notation:

expectation/mean of a R.V.

$$\mathbb{E}[X] \triangleq \sum_{x \in \Omega_X} x\, p(x) \quad \text{(for discrete R.V.)}$$

$$\int_{\Omega} x\, p(x)\, dx \quad (\text{" cts. "})$$

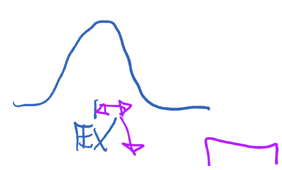
$\mathbb{E}[\cdot]$ is a linear operator $\Rightarrow$ $\mathbb{E}[\underbrace{a}_{\text{scalar}}\overbrace{X}^{\text{R.V.}} + \overbrace{Y}^{\text{R.V.}}] = a\,\mathbb{E}[X] + \mathbb{E}[Y]$

variance:

$$\mathrm{Var}[X] \triangleq \mathbb{E}\left[(X - \mathbb{E}[X])^2\right]$$

$$= \mathbb{E}X^2 - (\mathbb{E}X)^2$$

"measure of dispersion"



$\mathbb{E}X$ ↓
$\sigma \triangleq \sqrt{\mathrm{Var}[X]}$
"standard deviation"

parametric models:

↳ a family of distributions

$$\mathcal{P}_{\Theta} = \{ \underbrace{p(\cdot\,;\theta)}_{\text{possible pmf / pdfs depending on parameter } \theta} \mid \theta \in \Theta \}$$

set of valid parameters (Θ)

from context:

"support of distribution" i.e. Ω_X is usually fixed for all θ

examples: $\Omega_X = \{0,1\}$ for a coin flip

$\Omega_X = [0, \infty)$ for gamma dist. family

abuse of notation: $\{ p(x;\theta) \mid \theta \in \Theta \}$

"better" notation $\{ p(X;\theta) \mid \theta \in \Theta \}$

notation: $X \sim \mathrm{Bern}(\theta)$

"R.V. X is distributed as a Bernoulli dist. with parameter θ"

$p(x;\theta) = \mathrm{Bern}(x;\theta)$

parameter of dist.

variable for pmf

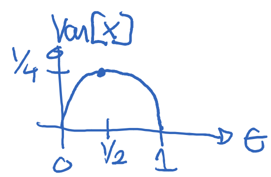
Bernoulli: prob. of a coin flip $\Omega_X = \{0,1\}$

$$P\{X=1 \mid \theta\} = \theta \quad \Rightarrow \Theta = [0,1]$$

$$p(x;\theta) = \theta^{x}(1-\theta)^{1-x} = \mathrm{Bern}(x;\theta)$$

$\mathbb{E}[X] = \theta$

$\mathrm{Var}[X] = \theta(1-\theta)$



$\mathrm{Var}[X;\theta]$
$\mathrm{Var}_\theta[X]$
$\mathrm{Var}[X \mid \theta]$

15h35

another example: $p(x; (\mu,\sigma^2)) = N(x \mid \mu, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)$

or $N(x; \mu, \sigma^2)$

binomial distribution: model n indep. coin flips

sum of n indep. $Bern(\theta)$ R.V.s

(note: is usually denoted by p)

let $\underbrace{X_i}_{} \overset{iid}{\sim} Bern(\theta)$ → independent & identically distributed

implicitly talking about $X_1, \ldots, X_n$

let $X \doteq \sum_{i=1}^{n} X_i$; then we have $X \sim Bin(n, \theta)$

"binomial with parameters n & θ"

$\Omega_X = \{0, 1, \ldots, n\}$

$$p(x;\theta) = \underbrace{\binom{n}{x}}_{} \underbrace{\theta^{x}(1-\theta)^{n-x}}_{} \quad \text{for } x \in \Omega_X = 0:n$$

# of ways to choose x elements out of n "n choose x"

$$\binom{n}{x} \triangleq \frac{n!}{x!(n-x)!}$$

$$\rightarrow \theta^{\sum_i x_i}(1-\theta)^{\sum_i (1-x_i)}$$

$$= \prod_{i=1}^{n} Bern(x_i;\theta) = p(x_1, \ldots, x_n)$$

mean: $X = \sum_i X_i$

$$\mathbb{E}[X] = \sum_i \mathbb{E}[X_i] = n\theta$$

similarly $Var[\sum_i X_i] \overset{\text{indep.}}{=} \sum_i Var[X_i] = n\theta(1-\theta)$

other distributions: $Poisson(\underset{\text{mean}}{\lambda})$ $\Omega_X = \{0, 1, \ldots\} = \mathbb{N}$ [count data]

Gaussian in 1D $N(\underset{\text{mean}}{\mu}, \underset{\text{variance}}{\sigma^2})$ $\Omega_X = \mathbb{R}$

gamma $\Gamma(\underset{\text{shape}}{\alpha}, \underset{\text{scale}}{\beta})$ $\Omega_X = \mathbb{R}^{+}$

other: Laplace, Cauchy, exponential, beta, Dirichlet, etc.

statistical concepts

prob. theory } well defined

epistemology

θ

Statistical concepts

cartoon

$$\underset{\text{model}}{\theta} \;\xrightarrow{\text{prob. theory}}\; \underset{\text{data}}{X}$$
} well defined

$$\underset{\text{data}}{X} \;\xrightarrow{\text{statistics}}\; \underset{\text{model}}{\theta}$$
} ill-defined "inverse problem"

epistemology of science!

example: model n indep. coin flips

prob. theory → prob. k heads in a row

statistics: I have observed k heads, $n-k$ tails, what is θ?

frequentist vs. Bayesian:

semantic of prob.: meaning of a prob.?

a) (traditional) frequentist semantic

$P\{X=x\}$ represents the limiting frequency of observing $X=x$

IF I could repeat ∞ # of i.i.d. experiments

b) Bayesian (subjective) semantic

$P\{X=x\}$ encodes an agent "belief" that $X=x$

laws of prob. characterize a "rational" way to combine "beliefs" and "evidence" [observations]

[→ has motivation in terms of gambling, utility/decision theory, etc.]

operationally:

Bayesian approach: ⊛ very simple philosophically:

- treat all uncertain quantities as R.V.
- i.e. encode all knowledge about the system ("beliefs") as a "prior" on probabilistic models and then use law of prob (and Bayes rule) to get answers!

frequentist justification:

- for a discrete R.V., suppose $P\{X=x\}=\theta$

$$\Rightarrow P\{X\neq x\}=1-\theta$$

$$B \triangleq \mathbb{1}\{X=x\} \qquad \Rightarrow B\sim \text{Bern}(\theta) \text{ R.V.}$$

indicator fct. $\mathbb{1}_A(u)=\begin{cases}1 & \text{if } u\in A\\ 0 & \text{o.w.}\end{cases}$

repeat i.i.d. experiments i.e. $B_i \overset{iid}{\sim} \text{Bern}(\theta)$

by L.L.N. (law of large numbers) $\frac{1}{n}\sum_{i=1}^{n} B_i \xrightarrow{a.s.} \mathbb{E}[B_1]=\theta$

limiting frequency

by CLT $\sqrt{n}\left(\frac{1}{n}\text{Bin}(n,\theta)-\theta\right) \xrightarrow{d} N(0,\theta(1-\theta))$