Lecture 3 - parametric models

Tuesday, September 12, 2023 2:34 PM

today:
- prob. review
- parametric models

Bayes rule

$\rightarrow$ invert conditioning

$$p(x|y) = \frac{p(y|x)p(x)}{p(y)}$$

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

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

"product rule"

example of conditioning: $P\{$ having cancer | tumor measurement = "-" $\}$

chain rule: $\rightarrow$ successive application of product rule

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

$$= \quad \quad \quad \quad \parallel \quad \quad p(x_{n-1}|x_{1:n-2})\,p(x_{1:n-2})$$

$$\boxed{p(x_1,\dots,x_n) = \prod_{i=1}^{n} \underbrace{p(x_i|x_{1:i-1})}}$$

convention: $1:0 = \emptyset$

$$p(x_1|x_\emptyset) = p(x_1)$$

always true

[can also choose any of $n!$ permutations]

can be simplified by making conditional independence assumptions

[for a directed graphical model]

$$p(x_i|x_{1:i-1}) = p(x_i|x_{\pi_i})$$

π_i: parents of i

conditional independence:

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

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

"factorization view"

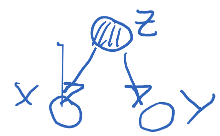
exercise to the reader: prove that $X \perp\!\!\!\perp Y|Z \Rightarrow p(x|y,z) = p(x|z)$

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

<u>example</u>: Z = indicator whether mother carries a genetic disease

X = daughter 1 has disease
Y = " 2 has "

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



here, X is <u>not</u> "marginally indep." of Y
(i.e. $X \not\perp\!\!\!\perp Y$) $\exists\, x,y$ s.t. $p(x,y) \neq p(x)p(y)$

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

<u>example of pairwise indep. R.V. that are <u>not</u> mutually independent</u>

X = coin flip as $\{0,1\}$
Y = another indep. coin flip
$X \perp\!\!\!\perp Y$

<u>define</u>: $Z \triangleq X \text{ xor } Y$
$Z \perp\!\!\!\perp Y$
$Z \perp\!\!\!\perp X$
$Z \not\perp\!\!\!\perp (X,Y)$

<u>other notation</u>:

<u>expectation / mean of</u> a R.V.

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

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

$\mathbb{E}[\cdot]$ is a linear operator $\Rightarrow$ $\mathbb{E}[aX + Y] = a\,\mathbb{E}[X] + \mathbb{E}[Y]$

R.V.'s

scalar

<u>variance</u>:

$$\mathrm{Var}[X] \triangleq \mathbb{E}[(X - \mathbb{E}[X])^2] \quad \text{"measure of dispersion"}$$

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

<u>cumulative dist. function</u> CDF of X: $F_X(t) = P\{X \leq t\}$ $\left(= \int_{-\infty}^{t} p(x)\,dx \right)$

$$\frac{d}{dt} F_X(t) \Big|_{t=x} = p(x) \longrightarrow \text{(for cts. R.V.)}$$

15h30

parametric models:

↳ a family of distributions

$$\mathcal{P}_{\Theta} = \{ \underbrace{p(\cdot;\theta)}_{\text{possible pmf's / pdf's}} \mid \theta \in \Theta \}$$

depend on parameter θ

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

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

__from context__

"support of ~~distribution~~ family" i.e. Ω_X

is __usually__ __fixed__ for all $\theta \in \Theta$

(but not always)

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

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

__notation:__ $X \sim \text{Bern}(\theta)$

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

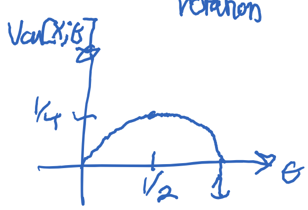
$$p(x;\theta) = \text{Bern}(\underset{\substack{\uparrow \\ \text{variable} \\ \text{for pmf}}}{x};\underset{\substack{\uparrow \\ \text{parameter} \\ \text{of dist.}}}{\theta}) \triangleq \theta^{x}(1-\theta)^{1-x} \qquad x \in \{0,1\}$$

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

$p(x=1) = \theta$ $\qquad$ Bayesian notation $P\{X=1 \mid \theta\} = \theta$

$E[X] = \theta$

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



__another example:__

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

or

$$\mathcal{N}(x \mid \mu,\sigma^2)$$

binomial dist.:

model n __indep.__ coin flips

sum n indep. $\text{Bern}(\theta)$ R.V.

let $X_i \overset{\text{iid}}{\sim} \text{Bern}(\theta)$ → (mutually) independent and identically distributed implicitly talking $X_1, \dots, X_n$ $\quad (X_i)_{i=1}$ iid.

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

"binomial dist. with parameter n & θ"

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

pmf : $p(x;\theta) = \underbrace{\binom{n}{x}}_{\substack{\text{\# of ways} \\ \text{to choose } x \\ \text{elements out of } n \text{ possibilities}}} \theta^{x}(1-\theta)^{n-x}$

"n choose x"

derivation

$\text{Bern}(x_i;\theta) = \theta^{x_i}(1-\theta)^{1-x_i}$

$p(x_1, \dots, x_n) \underset{\text{by indep}}{=} \prod_{i=1}^{n} \text{Bern}(x_i;\theta) = \theta^{\overbrace{\sum_i x_i}^{\triangleq x}}(1-\theta)^{n-\sum_i x_i}$

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

mean: $X = \sum_i X_i$

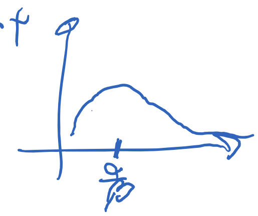
$\mathbb{E}[X] = \sum_i \mathbb{E}[X_i] = n\theta$

similarly, $\text{Var}\left(\sum_i X_i\right) \overset{\text{indep}}{=} \sum_i \text{Var}(X_i) = n\theta(1-\theta)$

<u>other distributions:</u> $\text{Poisson}(\underset{\text{mean}}{\lambda})$ $\quad \Omega_X = \{0, 1, \dots, \} = \mathbb{N}$ [count data]

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

gamma $\quad \Gamma(\underset{\text{shape}}{\alpha}, \underset{\text{inverse scale}}{\beta})$ aka rate $\quad \Omega_X = \mathbb{R}_+$



mean: $\frac{\alpha}{\beta}$ $\quad$ variance $\frac{\alpha}{\beta^2}$

other : Laplace, Cauchy, exponential dist. $\to \Gamma(1, \beta)$, beta dist. $\to \Omega_X = [0,1]$ is a Dirichlet on 2 elements