Lecture 3 - prob review / parametric models

Tuesday, September 13, 2022 2:59 PM

today: finish prob. review
parametric models

independence:

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

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

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

sidenote: $Z = (\underset{\text{cts.}}{X}, \underset{\text{discrete}}{Y})$

$P\{X \in \text{box}, Y = y\} = \int_{x \in \text{box}} p(\overset{\text{pdf part}}{x}, \underbrace{y}_{\text{pmf part}})\, dx$

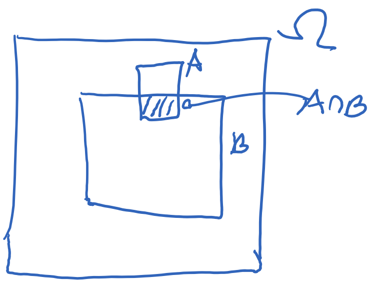
$P\{X \in \text{box}, Y \in \text{set}\} = \int_{x \in \text{box}} \left(\sum_{y \in \text{set}} p(x,y) \right) dx$

Conditioning:

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

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

normalization



event $E \subseteq \Omega$
$E \in \mathcal{E}$

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

for discrete R.V.:

define "conditional pmf"

$p(x|y) \triangleq P(X=x \mid Y=y) \triangleq \dfrac{P(\{X=x, Y=y\})}{P(\{Y=y\})}$ } joint pmf / } marginal pmf

$p(x|y) \underset{\text{"proportional to"}}{\propto} p(x,y)$ normalization cst. $\sum_x p(x,y) = p(y)$

for cts. R.V.: "conditional pdf" $p(x|y) \triangleq \dfrac{p(x,y)}{p(y)}$

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

Bayes rule:

→ 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: → successive application of product rule

$$p(x_1, \dots, x_n) = p(x_n | x_{1:n-1})\, \underbrace{p(x_{1:n-1})}_{p(x_{n-1}|x_{1:n-2})\,p(x_{1:n-2})}$$

$$= \quad \text{"}$$

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

always true

convention $1:0 = \emptyset$ $\quad p(x_1 | x_\emptyset) = p(x_1)$

→ can be simplified by making conditional independence assumptions [from directed graphical model]

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

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

conditional independence:

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

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

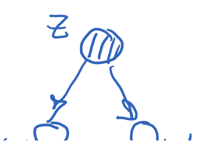
exercise to the reader: prove that $X \perp\!\!\!\perp Y \mid Z \Rightarrow p(x|y,z) = p(x|z)$

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

example: Z = indicator whether mother carries a genetic disease

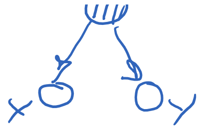
X = son 1 has disease $\qquad X \perp\!\!\!\perp Y \mid Z$

Y = son 2 " "



here, X is not "marginally independent" of Y

(i.e. $X \not\perp\!\!\!\perp Y$) $\quad \exists x,y$ s.t. $p(x,y) \neq p(x)p(y)$



here, X is not 'marginally indep't' of Y

(i.e. $X \not\perp\!\!\!\perp Y$) $\exists\, x, y$ s.t. $p(x,y) \neq p(x)p(y)$

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

16h05

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

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

Y = another indep. coin flip

$X \perp\!\!\!\perp Y$

<u>define</u> $Z \triangleq X \text{ xor } Y$

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

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

$Z \not\perp\!\!\!\perp (X,Y)$

<u>other notation</u> :

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

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

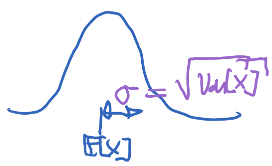
$$\int_\Omega x\, p(x)\, dx \qquad \text{(" cts. ")}$$

$\mathbb{E}[\cdot]$ is a linear operator $\Rightarrow$ $\mathbb{E}[\underset{\text{scalar}}{a} X + Y] = a\,\mathbb{E}[X] + \mathbb{E}[Y]$ (R.Vs.)

<u>variance</u> :

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

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



<u>cumulative distribution fct.</u> CDF of X : $F_X(t) = \mathbb{P}\{X \leq t\} \left(= \int_{-\infty}^{t} p(x)\,dx \right)$ [for a cts. R.V.]

$$\left.\frac{d}{dt} F_X(t)\right|_{t=x} = p(x) \quad \text{for a cts. R.V.}$$

<u>parametric models</u> :

↳ a family of distributions

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

<u>from context</u>

"support of distribution" i.e. Ω_X is usually fixed for all θ (but not always)

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

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

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

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

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

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

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

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

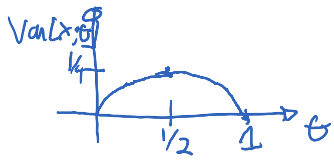
$p(x;\theta) = \mathrm{Bern}(\underbrace{x}_{\text{variable for pmf}};\ \overbrace{\theta}^{\text{parameter of dist.}}) = \theta^{x}(1-\theta)^{1-x} \qquad x \in \{0,1\}$

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

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

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

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



another example:

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

or

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

binomial dist.:

model n indep. coin flips

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

let $\underbrace{X_i}$ $\overset{\text{iid}}{\sim}$ $\mathrm{Bern}(\theta)$ $\quad$ iid $\rightarrow$ (mutually) independent and identically distributed

implicitly talking $\leftrightarrow$ $X_1, \dots, X_n$

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

"binomial dist. with parameter n & θ"

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

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

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

$\prod_{i=1}^{n} \mathrm{Bern}(x_i;\theta) = \theta^{\underbrace{\sum x_i}_{x}}(1-\theta)^{n-\underbrace{\sum x_i}_{x}}$

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

mean: $X = \sum_i X_i$

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

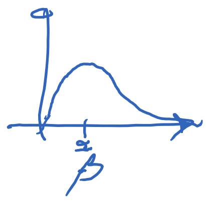
similarly, $\mathrm{Var}\left(\sum_i X_i\right) \overset{\text{indep.}}{=} \sum_i \mathrm{Var}(X_i) = n\theta(1-\theta)$

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

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

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

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



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

Dirichlet