

Lecture 3 - scribbles

Tuesday, September 8, 2020 14:26

today: • finish prob. review
• parametric models

joint, marginal, etc. (context: 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 dist.

can represent elementary events of $\Omega_{(X,Y)}$ as table:

	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 " " " " " " " " odd

cts. case: if X & Y are cts
↑
"and"

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

↑
joint pdf

marginal distribution (in the context of a joint dist.)

↳ dist. of a component of a random vector

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

"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" $\Leftrightarrow p(x_{1:n}) = \prod_{i=1}^n p(x_i) \quad \forall x_{1:n} \in \prod_{i=1}^n \Omega_{X_i}$

$$Z = (X, Y) \quad \begin{matrix} \uparrow & \uparrow \\ \text{cls.} & \text{discrete} \end{matrix} \quad P\{x \in \text{box}, Y=y\} = \int_{x \in \text{box}} p(x, y) dx$$

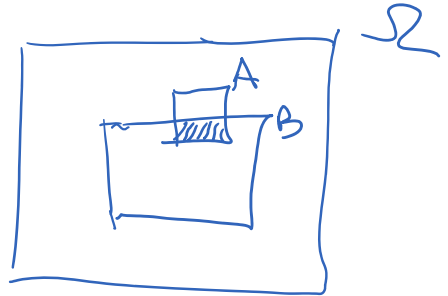
pdf part
pdf part

conditioning :

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

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

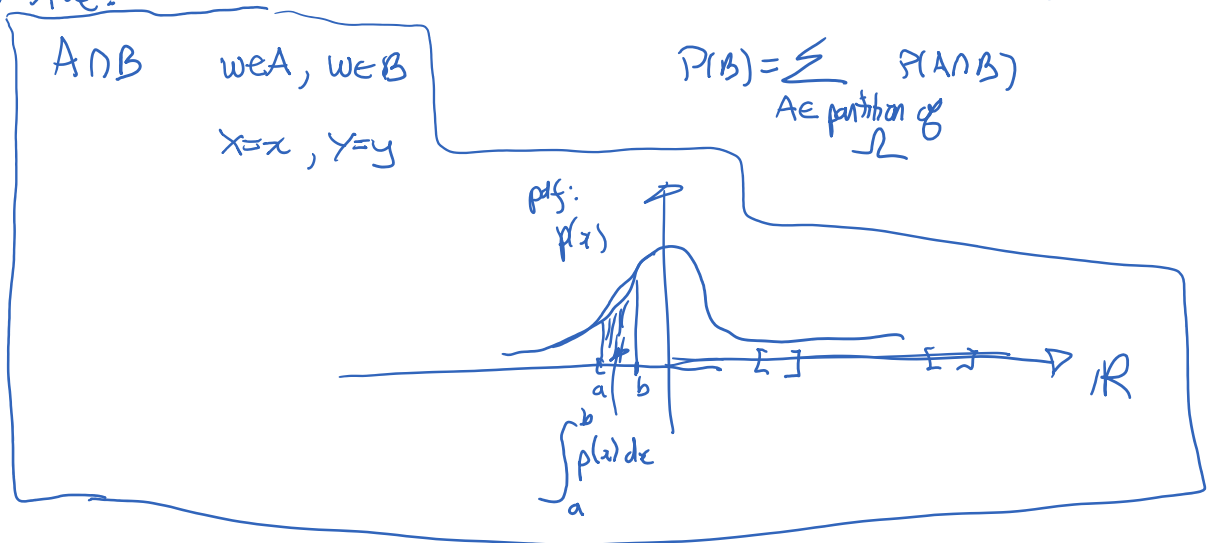
normalization



Aside:

$$A \cap B \quad \begin{matrix} w \in A, w \in B \\ X=x, Y=y \end{matrix}$$

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



- for discrete R.V.:

define "conditional pdf"

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

joint pdf
marginal pdf

$$p(x|y) \propto p(x, y)$$

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

- for cls. R.V.

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

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

[note independent $X \perp\!\!\!\perp Y$ $p(x,y) = \frac{p(x,y)}{p(y)} = \frac{p(x)p(y)}{p(y)} = p(x)$]

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

Note: the paradox about conditional probabilities that I was looking for was indeed (as Jacob Louis Hoover rightfully pointed out) the [Borel-Kolmogorov paradox](#) - have a look at it if you're curious!

Bayes rule:

→ invert conditioning

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

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

→ $p(x,y) = p(y|x)p(x) = p(x|y)p(y)$
 "product rule"

chain rule: → successive application of product rule

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

$$= \dots p(x_{n-1} | x_{1:n-2}) p(x_{1:n-2})$$

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

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

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

[can also choose any $n!$ permutation]

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

π_i parents in graph

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

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

example: $Z =$ indicator whether mother carries a genetic disease

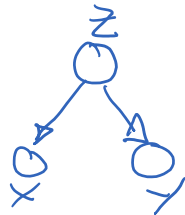
$X =$ son 1 has disease

$X \perp Y | Z$

$Y =$ " 2 " " "

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

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



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

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

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

$Y =$ ^{another} indep coin flip

$X \perp Y$

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

$Z \perp X$

$Z \perp Y$

~~$Z \perp (X,Y)$~~

other notation:

expectation/mean of a R.V.

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

$\int_{\Omega} x p(x) dx$ ("cfs. R.V.")

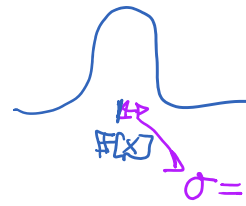
$E[\cdot]$ is a linear operator $\Rightarrow E[aX + bY] = aE[X] + bE[Y]$
↑
R.V.'s
↓
scales

variance:

$\text{Var}[X] \triangleq E[(X - E[X])^2]$

"measure dispersion"

$= E[X^2] - (E[X])^2$



"standard deviation"

parametric models ;

↳ a family of distributions

$$\mathcal{P}_{\Theta} = \{ p(\cdot; \theta) \mid \theta \in \Theta \}$$

possible pdf's / pmf's
depending on parameter θ

from context :

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

example ; $\Omega_X = \{0, 1\}$ for a coin flip

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

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

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

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

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

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

parameter of dist. θ
variable for pmf

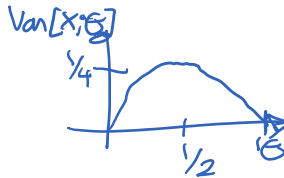
another example
 $N(x; \mu, \sigma^2)$

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

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

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

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



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