

today: • finish proba review
• parametric models



cdf: cumulative distribution function (cdf) $F_X(t) = P\{X \leq t\}$

if X is cts. R.V. $F_X(t) = \int_{-\infty}^t p(x) dx$

$\Rightarrow \frac{d}{dt} F_X(t) = p(t)$

joint, marginal, etc. (context: multivariate R.V.)

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

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

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

joint distribution

X as before
Y = die result
 $\in \{1, 2, 3, 4, 5, 6\}$

	X=0	X=1
Y=0	$\frac{1}{3}$	$\frac{1}{6}$
Y=1	$\frac{1}{6}$	$\frac{1}{3}$

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

	X=0	X=1
Y=0	○	○
Y=1	○	○

$p(1,1)$ (pointing to the bottom-right cell)

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

cts. case: if $X \& Y$ are cts "land"
 $P\{\text{box region}\} = \int_{\text{box}} p(x,y) dx dy$ (joint pdf)

marginal distribution (in the context of a joint dist)

\rightarrow 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 "

independence:

X is independent of $Y \Leftrightarrow p(x,y) = p(x)p(y) \quad \forall (x,y) \in \Omega_x \times \Omega_y$

notation: $X \perp Y$

$$p_{(X,Y)}(x,y) = p_X(x)p_Y(y)$$

in this notation $p(x,y) \neq p(y,x)$ are same thing

• R.V. X_1, \dots, X_n are "mutually independent" $\Leftrightarrow p(x_{1:n}) = \prod_{i=1}^n p(x_i) \quad \forall x_{1:n} \in \prod_{i=1}^n \Omega_{x_i}$

$Z = (X, Y)$
 cts. discrete

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

↑ event
↑ pmf part
↑ pmf part

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

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

$$\{X \in B\} \quad \{\omega: X(\omega) \in B\}$$

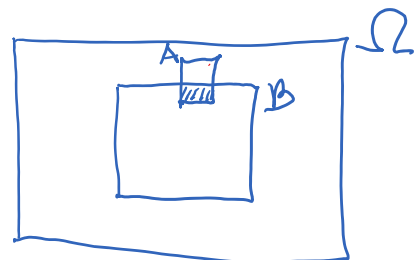
standard prob. / slab formulation

conditioning:

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

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

normalization



for discrete R.V.:

define "conditional pmf"

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

} joint pmf
} marginal pmf

alpha: \propto
 proportional to
 infinity: ∞

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

↑
↓
 $p(y)$

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

for a cts. R.V.:

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

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

see Borel-Kolmogorov paradox for weirdness of conditioning on prob. 0 events...

[Borel-Kolmogorov paradox](#) on wikipedia

$$[\text{note: independent } X \perp\!\!\!\perp Y \quad p(x|y) = \frac{p(x,y)}{p(y)} \stackrel{\text{by def}}{=} \frac{p(x)p(y)}{p(y)} = p(x)]$$

16h05

TA OH will Thursday 1-2pm

example of conditioning: $P(\text{having cancer} | \text{tumor measurement} = "x")$

Bayes rule

→ invert conditioning

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

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

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

"product rule"

chain rule: → successive application of product rule

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

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

always true

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

[can also choose any $n!$ permutation]

can be simplified when making conditional independence assumptions (from directed graphical models)

$$p(x_i | x_{1:i-1}) = p(x_i | \underbrace{x_{\pi(i)}}_{\text{parents in graph}})$$

conditional independence:

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

$$\Leftrightarrow p(x,y|z) = p(x|z)p(y|z) \quad \forall x \in \Omega_x, y \in \Omega_y$$

$$\bar{z} \in \Omega_z \text{ st. } p(\bar{z}) \neq 0$$

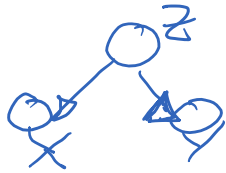
exercise to the reader: prove that if $X \perp\!\!\!\perp Y | Z$

$$\text{then } p(x|y,z) = p(x|z)$$

$$[\text{conditional analog of } X \perp\!\!\!\perp Y \Rightarrow p(z|y) = p(z)]$$

example: Z = indicator whether mother carries a genetic disease
 X = daughter 1 has disease
 Y = " 2 " "

$$X \perp\!\!\!\perp Y | Z$$



here, X is not "marginally indep" of Y

(i.e. $X \not\perp\!\!\!\perp Y$)

$$\exists x,y \text{ st. } p(x,y) \neq p(x)p(y)$$

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

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

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

Y = another indep coin flip.

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

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

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

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

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

other notation:

expectation / mean of a R.V.

$$\mathbb{E}[X] \triangleq \sum_{x \in \Omega_X} x p(x)$$

(for discrete R.V.)

$$\int_{\Omega} x p(x) dx$$

(" cts. R.V.)

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

\swarrow R.V.'s
 \searrow scalar

variance:

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

"measure dispersion"

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



parametric models:

↳ a family of distributions

see outside

↳ a family of distributions

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

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

from context:

"support of distribution" is Ω_X
is usually fixed for all θ

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

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

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

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

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

$$p(x; \theta) = \text{Bern}(x; \theta) = \theta^x (1-\theta)^{1-x} \quad x \in \{0, 1\}$$

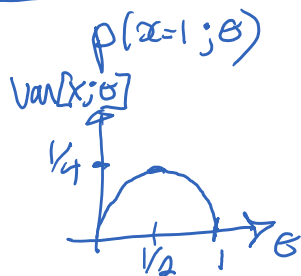
parameter of dist. θ
variable for pmf x

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

Bayesian notation $\mathbb{P}\{X=1 \mid \theta\} = \theta \Rightarrow \mathcal{H} = [0, 1]$

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

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



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

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