

Lecture 2 - scribbles

Tuesday, September 6, 2016
12:31

today: probability review
statistics

William → sunke

Probability review

why? → principled framework to model uncertainty

- Sources of uncertainty:
- a) intrinsic (Q.M.)
 - b) reasoning about future events
 - c) partial information about complex phenomenon
 - throwing a dice [initial conditions?]
 - object recognition model [complex?]

notation: X_1, X_2, X_3 or X_i, Y_i, Z

→ random variables (usually real-valued)

x_1, x_2, x_3 x, y, z

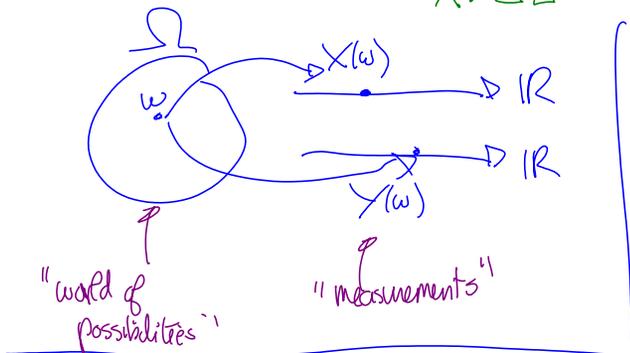
→ their realizations

$\{X_1 = x_1\}$ represents event
that R.V. X_1 takes value x_1

formally:

Ω → sample space of "elementary events"
 $\omega_1, \omega_2, \omega_3, \dots$ (assume countable)

then a random variable is a (measurable) mapping
 $X: \Omega \rightarrow \mathbb{R}$



probability distribution P

is a mapping $P: \mathcal{E} \rightarrow [0, 1]$

$\mathcal{E} \triangleq$ set of all subsets of Ω

↳ set of "events"

↳ " σ -field" [needed when Ω is uncountable]

Kolmogorov axioms $\left\{ \begin{array}{l} \text{S.t. I) } P(E) \geq 0 \quad \forall E \in \mathcal{E} \\ \text{II) } P(\Omega) = 1 \\ \text{III) } P(\bigcup_{i=1}^n E_i) = \sum_{i=1}^n P(E_i) \text{ when } E_i \text{'s are } \underline{\text{disjoint}} \end{array} \right.$

prob on Ω induces a prob. on dist. on image $X \quad \Omega_X \triangleq X(\Omega)$

e.g. "event" $\{x\}$ in Ω_X , get $P_X(\{x\}) = P\{\omega : X(\omega) = x\}$

$$= P\{X^{-1}(\{x\})\}$$

$$= P\{X = x\}$$

$$= p(x) \quad (?)$$

notation $X^{-1}(A) \triangleq \{\omega : X(\omega) \in A\}$ shorthand: ω (even onto) image of $X = \{x : \exists \omega \in \Omega \text{ s.t. } X(\omega) = x\}$

Example: $\Omega = \{1, 2, \dots, 6\}$ dice roll

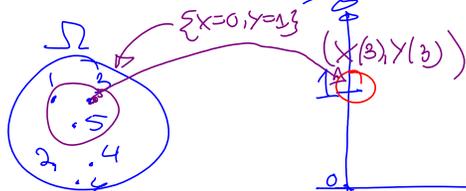
e.g. X looks / "measures" whether dice result is even
 Y " " " " " " is odd

$$X = \mathbb{1}_{\{2, 4, 6\}}$$

by defn: $\mathbb{1}_A(\omega) = \begin{cases} 1 & \text{if } \omega \in A \\ 0 & \text{o.w.} \end{cases}$

indicator function on A

similarly $Y = \mathbb{1}_{\{1, 3, 5\}}$



$(X(\omega), Y(\omega))$

joint distribution on $(x, y) \in \Omega_X \times \Omega_Y$

← "random vector"

= "vector valued R.V."

$$P_{X,Y}\{X=x, Y=y\}$$

means "and" \Rightarrow intersection of events

is the generalized sense

$$= P(X^{-1}(\{x\}) \cap Y^{-1}(\{y\}))$$

can represent as table:

	$X=0$	$X=1$	
$Y=0$	0	$\frac{1}{2}$	$P\{X=1, Y=0\}$ $= P\{2, 4, 6\}$ $= \sum_{\omega \in \{2, 4, 6\}} p(\omega) \rightarrow \frac{1}{6}$
$Y=1$	$\frac{1}{2}$	0	

marginal distribution: (in context of a joint dist.) distribution on components of the random vector

$$P\{X=x\} = \sum_y P\{X=x, Y=y\} \quad \text{"sum rule"}$$

- $y \in \Omega_X$ -

types of R.V. :

discrete R.V. $\rightarrow \Omega_X$ is countable

its distribution P_X (on Ω_X) is fully defined

by probability mass function "pmf." $P_X(\{X=x\})$ for $x \in \Omega_X$

shorthand $P_X(x)$ ($= P_X(\{X=x\})$)

or even $p(x)$

! "typed evaluation"

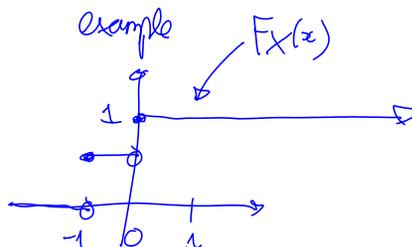
$p(x) \neq p(y)$ even if " $x=y$ " (!!!)
means " $P_X(x) \neq P_X(y)$ "

other R.V. basics :

for $\Omega_X \in \mathbb{R}$, prob. dist. P_X is fully characterized by its cumulative distribution function (cdf)

$$F_X(x) \triangleq P_X\{X \leq x\}$$

- properties:
- F_X is non-decreasing
 - $\lim_{x \rightarrow -\infty} F_X(x) = 0$
 - $\lim_{x \rightarrow +\infty} F_X(x) = 1$



(for discrete R.V., cdf is piecewise constant)

* for a "continuous R.V." , the cdf is "absolutely cts."

(\Leftrightarrow differentiable almost everywhere)

and \exists function $f(x)$
s.t. $F_X(x) = \int_{-\infty}^x f(u) du$

probability density function

pdf for a cts R.V.

is the analog of the pmf for a discrete R.V.

replace \rightarrow (for discrete R.V.)

$$\frac{d}{dx} F_X(x) = f(x) \Leftrightarrow \text{true where } f \text{ is cts}$$

with \int (for cts. R.V.)
 e.g. $\sum_{x \in \Omega_x} p(x) = 1$ (pmf) (discrete R.V.)

$\int_{\Omega_x} p(x) dx = 1$ (pdf) (for a cts. R.V.)

note $p(x)$ can be bigger than 1 for a pdf

e.g. uniform distribution $[0, \frac{1}{2}]$
 then $p(x) = \begin{cases} 2 & \text{for } x \in [0, \frac{1}{2}] \\ 0 & \text{otherwise} \end{cases}$

* Expectation / mean of a R.V.

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

$E[\cdot]$ is a linear operator: $E[aX + Y] = aE[X] + E[Y]$

variance:

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

does not always exist \rightarrow
 e.g. Cauchy distribution

"measure of dispersion"

independence

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

notation $X \perp Y$

"mutual independence" for R.V. $X_1, \dots, X_n \iff p(x_1, \dots, x_n) = \prod_{i=1}^n p(x_i)$

conditioning

for events A & B , suppose that $p(B) \neq 0$

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

for R.V., define $p(X=x | Y=y) \triangleq \frac{p(X=x, Y=y)}{p(Y=y)}$

$$\sum_x p(x=y) = p(x=y) \quad \text{Normalization constant}$$

example: "P(having cancer | tumor measurement)"

by def'n we get product rule $p(x, y) = p(x|y)p(y)$

subtle pt. $p(x|y)$ is undefined when $p(y)=0$

⊗ Bayes rule → inverting the conditioning

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

Chain rule → successive application of product rule

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

= ...

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

(always true)

↘ can simplify this using conditional indep assumptions [directed graph model]

conditional indep: X is cond. indep of Y given Z

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

$$\Leftrightarrow p(x, y|z) = p(x|z)p(y|z) \quad \forall x \in \Omega_x, y \in \Omega_y, z \in \Omega_z \text{ s.t. } p(z) > 0$$

example Z = mother carries genetic disease on X chromosome

X = son 1 has disease

Y = son 2 " "

exercise to the reader: prove that $p(x|y, z) = p(x|z)$

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