Lecture 2 - scribbles
Friday, September 8, 2017
13:33

Today's probability theory review

(aside: RVM 😞)

why? -> principled framework to model uncertainty

sources of uncertainty

1) intrinsic uncertainty -> quantum mechanics

2) partial information / observation
   e.g. rolling a dice -> don't know exactly the initial conditions

3) incomplete modeling of complex phenomenon
   e.g. "most birds can fly"
   -> simplicity of rule is advantage but also yields uncertainty

   * object recognition model

notation: $X_1, X_2, X_3$ or $X, Y, Z$

$X_1, X_2, X_3$ random variables (usually real-valued)

$X_1, X_2, X_3$ or $Y, Z$

$\omega, \gamma, \zeta$ their realizations

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

examples: rolling a dice $\Omega = \{1, 2, 3, 4, 5, 6\}$ sample space of "elementary events"
example: rolling a die \( \Omega = \{1, 2, 3, 4, 5, 6\} \) sample space of "elementary events"

R.V. \( X = \{1, 2, 3, 4, 5, 6\} \)

\( 1_A \) \( = \) indicator function on set \( A \)

\[ 1_A(w) = \begin{cases} 1 & \text{if } w \in A \\ 0 & \text{if } w \notin A \end{cases} \]

\( \Omega \) \( = \) \{1,2,3,4,5,6\} \( \Rightarrow \) event that dice output was even here

Formally:

\( \Omega \) is a sample space of "elementary events"

\( \{w_1, w_2, w_3, \ldots\} \) (assume countable for now)

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

```
\( \Omega = \{1, 2, 3, 4, 5, 6\} \)

\( X = \{1, 2, 4, 6\}, \) "even"

\( Y = \{1, 3, 5\}, \) "odd"
```

\( \Omega \) \( = \) \{1,2,3,4,5,6\} \( \Rightarrow \) world of elementary possibilities

a probability distribution \( P \)

is a mapping \( P: \Omega \rightarrow [0, 1] \)

\( E \) \( = \) set of all subsets of \( \Omega \)

\( F \) \( = \) set of "events" (is a field" in measure theory needed when \( \Omega \) is uncountable)
a probability distribution \( P \) is a mapping \( P : \Omega \rightarrow [0, 1] \)

\( \mathcal{E} \) is the set of all subsets of \( \Omega \).

\( \mathcal{F} \) is the set of "events" ("sigma-field" in measure theory). Needed when \( \Omega \) is uncountable.

Which satisfies the following properties:

1. \( P(E) \geq 0 \) \( \forall E \in \mathcal{E} \)
2. \( P(\emptyset) = 1 \)
3. \( P\left( \bigcup_{i=1}^{\infty} E_i \right) = \sum_{i=1}^{\infty} P(E_i) \) when \( E_i \)s are disjoint.

Prob. on \( \Omega \) induces a prob. dist. on image of \( X \):

\( X(\Omega) = \{ y : \exists w \in A \text{ s.t. } X(w) = y \} \)

\( X^{-1}(B) = \{ w : X(w) \in B \} \)

\( X^{-1}(\emptyset) = \{ w : X(w) \in \emptyset \} \)

Universe was \( X^{-1}(\emptyset) \) is set of \( w \)s s.t. \( X(w) = \emptyset \)

(e.g. "event" \( x=3 \) in \( X \), get \( X^{-1}(\{3\}) = \{ w : X(w) = 3 \} \))
Recall \( \mathbb{E}(X^3) \) represents both the event \( x \in \Omega \) and \( \forall x \in \Omega \).

Let even \( \forall \omega \in \Omega \): \( X(\omega) = x \) \( \in \mathbb{E} \).

Back to own example (\( X \)): Let

\[
\begin{array}{c|c|c|c}
\omega & X(\omega) = 0 & X(\omega) = 1 \\
--- & --- & --- \\
\Omega & 0 & 1 & \text{(Intersection of events)}
\end{array}
\]

Joint distribution on \((x, y) \in \Omega_x \times \Omega_y\)

\[
P_{X,Y} \{ X=x, Y=y \}
\]

"Random vector" \( = \) "vector valued R.U." in the generalized sense

\[= \mathbb{P}(X \in 3 \mathbb{E}) \cap Y = (y, y, y) \]

can represent events of \( \Omega_x \times \Omega_y \) as like:

\[
\begin{array}{c|c|c|c}
\omega & X=0 & X=1 \\
--- & --- & --- \\
\Omega & 0 & 1 & \frac{1}{2} \\
Y=1 & \frac{1}{2} & 0 & \frac{1}{2}
\end{array}
\]

\[
P_{X,Y}=P(1,1,1) = \mathbb{P}(1,1,1,3) = \mathbb{P}(1,1,1,3) = \frac{1}{2}
\]
Marginal Distribution (in the context of the joint)

- Distribution on the components of a random vector

\[ P(\mathbf{X} = \mathbf{x}) = \sum_{y \in \mathcal{Y}} P(\mathbf{X} = \mathbf{x}, Y = y) \]

This summation procedure is called "marginalization".

From \( P_{X,Y} \) distribution for \((X,Y)\)

\( P_X \) "marginal distribution" \(\approx\) distribution for \(X\) by itself

Other R.V. basics:

- Types of R.V.:
  - "Discrete R.V." \(\Rightarrow\) \(\mathcal{S}_x\) is countable
  - "Continuous R.V." \(\Rightarrow\) \(\mathcal{S}_x\) is uncountable + \(f\) density

  - Its distribution \(P_X\) (on \(\mathcal{S}_x\)) is fully defined

  - By probability mass fn. "pmf" \(P(\mathbf{X} = \mathbf{x})\) for \(x \in \mathcal{S}_x\)

  - Shorthand: \(p_X(x)\)

  - Or even \(f_X(x)\) (?)

For any (valid) R.V., pdf distribution \(P_X\) is fully characterized

by its cumulative distribution function (CDF): \(F_X(x) = P(\mathbf{X} = \mathbf{x})\)

\(\mathbb{R} \to [0,1]\)

- Properties:
  - \(F_X\) is non-decreasing
  - \(\lim_{x \to -\infty} F_X(x) = 0\)

\[ F(x) \]

\[ \Phi \]

\(D(\mathbf{x} = \mathbf{A})\)
\[ F_X(x) = \begin{cases} 
0 \quad &x < 0 \\
\frac{x^n}{n!} \quad &0 \leq x \leq 1 \\
1 \quad &x > 1 
\end{cases} \]

*Continuous from the right*

(for a discrete R.V., the cdf is piecewise constant, and the jumps give values of pmf)

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

\[ \left( \Rightarrow \text{ differentiable almost everywhere} \right) \]

and a function \( f(x) \) s.t.
\[
F_X(x) = \int_{-\infty}^{x} f(t) \, dt
\]

\[ \frac{d}{dx} F_X(x) = f(x) \quad \text{when } x \text{ is a} \]

continuity pt. of \( f \)

*Probability density function (pdf)*

*pdf* for a cts. R.V. is the analog of the pmf for a discrete R.V.

with \( \sum \) (for cts. R.V.) replacing \( \Xi \)

\[ \sum_{x} p(x) = 1 \quad \text{for a discrete R.V.} \]

\[ \int_{-\infty}^{\infty} f(x) \, dx = 1 \quad \text{(all \( f(x) \) in ML)} \]

Note: \( f(x) \) (pdf) can be

began from \( 1 \)

for a pdf

e.g. uniform dist. on \([a, b] \)

then \( f(x) = \begin{cases} 
\frac{1}{b-a} & \text{for } x \in [a, b] \\
0 & \text{otherwise}
\end{cases} \)