today: probability review

statistical

William - sunke

Probability review

why? - principle framework to model uncertainty

Sources of uncertainty:

a) intrinsic (Q, M)
b) reasoning about future events
c) partial information about complex phenomenon

- throwing a die [initial conditions?]
- object recognition model [complex?]

notation: $X_1, X_2, X_3$ or $X_1, X_2, X_3$

→ random variables (usually real-valued)

$x_1, x_2, x_3$ $\rightarrow$ their realizations

$\exists X_1 = x_1 \text{ represents event that r.v. } X_1 \text{ takes value } x_1$

generally:

$\Omega$ - sample space of "elementary events"

$\omega_1, \omega_2, \omega_3, \ldots$ (assumed countable)

then a random variable is a (measurable) mapping

$X: \Omega \rightarrow \mathbb{R}$

probability distribution $P$

is a mapping $P: 2^\Omega \rightarrow [0, 1]$

$\mathcal{E} \triangleq \{ \text{set of all subsets of } \Omega \}$

$\mathcal{F} \triangleq \{ \text{set of "events"} \}$

$\mathcal{F} \triangleq \sigma-$field created when $\Omega$ is uncountable
Kolmogorov axioms

1. \( \sigma \)-additivity
2. \( \mathbb{P}(\emptyset) = 0 \)
3. \( \mathbb{P}(\Omega) = 1 \)

\( \mathbb{P}(A) = \sum \mathbb{P}(A_i) \) when \( A_i \) are disjoint

If \( \Omega \) induces a prob. on dist. on image \( X \), get \( P_X(x; x) = P \{ w : X(w) = x \} \)

\( \text{short hand: } \max \{ \text{even odds} \} = P(x) \)

Relation \( X^{-1}(A) = \{ w : X(w) \in A \} \) image of \( X = \{ x : \exists \omega \in \Omega, X(\omega) = x \} \)

\[ X^{-1}(A) = \{ w : X(w) \in A \} \]

Example:
\[ \Omega = \{1, 2, \ldots, 6\} \text{ dice roll} \]

\( X \) looks/"measures" whether dice result is even/odd.

\( X = \begin{cases} 1 & \text{if } 2, 4, 6 \\ 0 & \text{otherwise} \end{cases} \)

\( Y = \begin{cases} 1 & \text{if } 1, 3, 5 \\ 0 & \text{otherwise} \end{cases} \)

\[ \{ x \in \Omega : x = 1 \} \]

Joint distribution on \((X, Y) \in \Omega_X \times \Omega_Y\) 

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

\[ P \{ X \in A, Y \in B \} \]

Can represent as table:

<table>
<thead>
<tr>
<th>( X = 0 )</th>
<th>( X = 1 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>( Y = 0 )</td>
<td>( 0 ) ( 0 )</td>
</tr>
<tr>
<td>( Y = 1 )</td>
<td>( 0 ) ( 1 )</td>
</tr>
</tbody>
</table>

Marginal distribution (in context of a joint dist.)

\[ P_{X=1} = \sum P_{X=1, Y=1} \]

\[ P_{X=2} = \sum P_{X=2, Y=1} + P_{X=2, Y=3} \]

"Sum rule"
Types of R.V.: 

- **Discrete R.V.**: \( \Omega \times \text{is countable} \)

  Its distribution \( P_X \) (on \( \Omega \)) is fully defined by probability mass function (p.m.f.): \( P_X(\xi=x^2) \) for \( x \in \Omega \)

  Shorthand \( P_X(x) \) \( (= P_X(\xi=x^2)) \)

  Or even \( p(x) \)

  - \( \delta \) is a discrete r.v.

  \[
p(x) \neq p(y) \text{ even if } x = y \quad (\forall x, y)
\]

  \[
  F_X(x) \neq p(y)
\]

  Which means \( F_X(x) \neq p(y) \) even if \( x = y \) (\( \forall x, y \))

  \[
  F_X(x) \neq p(y)
\]

Other R.V. basics:

- For \( \Omega \times \subseteq \mathbb{R} \), prob. dist. \( P_X \) is fully characterized by its **cumulative distribution function** (c.d.f.)

  \[
  F_X(x) = P_X(\xi \leq x^2)
\]

  Properties:
  - \( F_X \) is non-decreasing
  - \( \lim_{x \to -\infty} F_X(x) = 0 \)
  - \( \lim_{x \to \infty} F_X(x) = 1 \)

  Example:

  \[
  F_X(x)
\]

  \[
  F_X(x)
\]

  (For discrete R.V., c.d.f. is piecewise constant)

* for a "continuous R.V.", the c.d.f. is "absolutely cts."

Pdf for a cts R.V.

- Is the analog of the p.m.f. for a discrete R.V.

  Replace \( \sum \) (for discrete R.V.)

  \[
  f_X(x) = \frac{d}{dx} F_X(x)
\]

  Probability density function

\( \Rightarrow \text{differentiable almost everywhere} \)

\( \Rightarrow \) function \( f(x) \)

\( \forall x \). \( F_X(x) = \int_{-\infty}^{x} f(y) \quad \text{d}y \)
with \( \sum_{x} (f(x) = 1) \) for a s.r.v.

e.g. \( \mathbb{E}[X] = \sum_{x \in \Omega_X} x f(x) = \int_{\Omega_X} x f(x) dx \) for a c.r.v.

\( \mathbb{E}[X] \) is a linear operator:
\[ \mathbb{E}[aX + b] = a \mathbb{E}[X] + b \mathbb{E}[X] \]

\[ \text{variance:} \]
\[ \text{Var}[X] = \mathbb{E}[(X - \mathbb{E}[X])^2] = \mathbb{E}[X^2] - (\mathbb{E}[X])^2 \]

\text{independence}

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

\text{relation} \( X \perp Y \)

"mutual independence" for r.v. \( X_1, \ldots, X_n \) \( \iff \) \( p(x_1, \ldots, x_n) = \prod_{i=1}^{n} p(x_i) \)

\text{conditioning}

\( A \cap B \) suppose that \( p(B) \neq 0 \)

Then define \( p(A | B) = \frac{p(A \cap B)}{p(B)} \)

\text{for r.v., define} \( p(x | y) = \frac{p(x, y)}{p(y)} \)
\( \frac{p(x=x, y=y)}{x} = p(x=x, y=y) \)

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

By definition we get the product rule: \( p(x, y) = p(x | y) p(y) \)

Notice that \( p(x | y) \) is undefined when \( p(y) = 0 \).

Bayes rule \( \Rightarrow \) Inverting the conditioning

\[
p(x | y) = \frac{p(y | x) p(x)}{p(y)}
\]

Chain rule \( \Rightarrow \) Successive application of product rule

\[
p(x_1, \ldots, x_n) = p(x_{1:n-1}) p(x_n | x_{1:n-1})
\]

\[
= \ldots
\]

\[
= \prod_{i=1}^{n} p(x_i | x_1, \ldots, x_{i-1})
\]

(always true) \( \checkmark \)

Can simplify this using conditional independence assumptions [directed graph model].

Conditional independence: \( X \) is conditionally independent of \( Y \) given \( Z \)

\[
X \perp Y | Z
\]

\[
p(x, y | z) = p(x | z) p(y | z)
\]

Example \( Z = \) mother carries genetic disease on X chromosome

\( X = \) son is carrier

\( Y = \) son has disease

Exercise to the reader: Prove that \( p(x | y, z) = p(x | z) \) when \( X \perp Y | Z \).