Probabilites

3 April 2017

Table of Contents I

Dice rolling game

Rolling a die (with numbers $1, ..., 6$).
Chance of getting a 5 $=$?

Dice rolling game

Rolling a die (with numbers $1, \ldots, 6$).
Chance of getting a $5 = ?$
$1/6$
Chance of getting a 5 or $4 = ?$

Dice rolling game

Rolling a die (with numbers $1, ..., 6$).
Chance of getting a $5 = ?$
$1/6$
Chance of getting a 5 or $4 = ?$
$2/6$

Dice rolling game

Rolling a die (with numbers $1, ..., 6$).
Chance of getting a 5 $=$?
$1/6$
Chance of getting a 5 or 4 $=$?
$2/6$

Probability $\approx$ a degree of confidence that an **outcome** or a number of outcomes (called **event**) will occur.

Probability Space

Probability space (a.k.a Probability triple) $(\Omega, \mathcal{F}, P)$:

- **Outcome space** (or sample space), denoted Ω (read "Omega") : the set of all possible outcomes[1].
    - roll a die: $\Omega = \{1, 2, 3, 4, 5, 6\}$. flip a coin: $\Omega = \{Head, Tail\}$.
- **σ-Field** (read "sigma-field", a set of events), denoted $\mathcal{F}$: An event ($\alpha \in \mathcal{F}$) is a set of outcomes.
    - roll a die to get 1: $\alpha = \{1\}$;
    - to get 1 or 3: $\alpha = \{1, 3\}$
    - roll a die to get an even number: $\alpha = \{2, 4, 6\}$
- **Probability measure** P: the assignment of probabilities to the events; *i.e.* a function returning an event's probability.

[1] of the problem that you are considering

Probability measure

Probability measure (or distribution) P over $(\Omega, \mathcal{F})$: a function from $\mathcal{F}$ (events) to $[0,1]$ (range of probabilities), such that,

- $P(\alpha) \geq 0$ for all $\alpha \in \mathcal{F}$
- $P(\Omega) = 1$, $P(\emptyset) = 0$.
- For $\alpha, \beta \in \mathcal{F}$, $P(\alpha \cup \beta) = P(\alpha) + P(\beta) - P(\alpha \cap \beta)$

Interpretations of Probability

- Frequentist Probability: $P(\alpha) =$ frequencies of the event. *i.e.* fraction of times the event occurs if we repeat the experiment indefinitely.
  - A die roll: $P(\alpha) = 0.5$, for $\alpha = \{2, 4, 6\}$ means if we repeatedly roll this die and record the outcome, then the fraction of times the outcomes in α will occur is 0.5.

Interpretations of Probability

- Frequentist Probability: $P(\alpha) =$ frequencies of the event. *i.e.* fraction of times the event occurs if we repeat the experiment indefinitely.
  - A die roll: $P(\alpha) = 0.5$, for $\alpha = \{2, 4, 6\}$ means if we repeatedly roll this die and record the outcome, then the fraction of times the outcomes in α will occur is 0.5.
  - Problem: non-repeatable event *e.g.* "it will rain tomorrow morning" (tmr morning happens exactly once, can't repeat).
- Subjective Probability: $P(\alpha) =$ one's own degree of belief that the event α will occur.

Conditional probability

Event α: "students with grade A"
Event β: "students with high intelligence"
Event $\alpha \cap \beta$: "students with grade A and high intelligence"

Conditional probability

Event α: "students with grade A"
Event β: "students with high intelligence"
Event $\alpha \cap \beta$: "students with grade A and high intelligence"

Question: how do we update the our beliefs given new evidence?

Conditional probability

Event α: "students with grade A"
Event β: "students with high intelligence"
Event $\alpha \cap \beta$: "students with grade A and high intelligence"

Question: how do we update the our beliefs given new evidence?
e.g. suppose we learn that a student has received the grade A, what does that tell us about the person's intelligence?

Conditional probability

Event α: "students with grade A"
Event β: "students with high intelligence"
Event $\alpha \cap \beta$: "students with grade A and high intelligence"

Question: how do we update the our beliefs given new evidence?
e.g. suppose we learn that a student has received the grade A, what does that tell us about the person's intelligence?

Answer: Conditional probability.
Conditional probability of β given α is defined as

$$P(\beta|\alpha) = \frac{P(\alpha \cap \beta)}{P(\alpha)}$$

Chain rule and Bayes' rule

- Chain rule: $P(\alpha \cap \beta) = P(\alpha)P(\beta|\alpha)$
  More generally,
  $P(\alpha_1 \cap ... \cap \alpha_k) = P(\alpha_1)P(\alpha_2|\alpha_1) \cdots P(\alpha_k|\alpha_1 \cap ... \cap \alpha_{k-1})$
- Bayes' rule:

$$P(\alpha|\beta) = \frac{P(\beta|\alpha)P(\alpha)}{P(\beta)}$$

Random Variables

Assigning probabilities to events is intuitive.

Assigning probabilities to attributes (of the outcome) taking various values might be more convenient.

- a patient's attributes such "Age", "Gender" and "Smoking history" ...
  "Age = 10", "Age = 50", ..., "Gender = male","Gender = female"
- a student's attributes "Grade", "Intelligence", "Gender" ...

$P(Grade = A)$ = the probability that a student gets a grade of A.

Random Variables

Random Variable[2] can take different types of values *e.g.* discrete or continuous.

- $Val(X)$: the set of values that X can take
- x: a value $x \in Val(X)$

Shorthand notation:

- $P(x)$ short for $P(X = x)$
- $\sum_x P(x)$ shorthand for $\sum_{x \in Val(X)} P(X = x)$

$$\sum_x P(x) = 1$$

[2]formal definition is omitted

Example

P(Grade, Intelligence).
Grade $\in \{A, B, C\}$
Intelligence $\in \{high, low\}$.

P(Grade = B, Intelligence = high) = ?
P(Grade = B) =?

		Intelligence		
		low	*high*	
Grade	*A*	0.07	0.18	0.25
	B	0.28	0.09	0.37
	C	0.35	0.03	0.38
		0.7	0.3	1

Marginal and Conditional distribution

Distributions:

- Marginal distribution $P(X) = \sum_{y \in Val(Y)} P(X, Y = y)$ or shorthand as $P(x) = \sum_y P(x, y)$
- Conditional distribution $P(X|Y) = \frac{P(X,Y)}{P(Y)}$

Rules for events carry over for random variables:

- Chain rule: $P(X, Y) = P(X)P(Y|X)$
- Bayes' rule: $P(X|Y) = \frac{P(Y|X)P(X)}{P(Y)}$

Independence and conditional independence

Independences give factorisation.

- Independence
  $X \perp\!\!\!\perp Y \Leftrightarrow P(X,Y) = P(X)P(Y)$
  - Extension: $X \perp\!\!\!\perp Y, Z$ means $X \perp\!\!\!\perp H$ where $H = (Y, Z)$.
    $\Leftrightarrow P(X,Y,Z) = P(X)P(Y,Z)$
- Conditional Independence
  $X \perp\!\!\!\perp Y|Z \Leftrightarrow P(X,Y|Z) = P(X|Z)P(Y|Z)$
  - Independence: $X \perp\!\!\!\perp Y$ can be considered as $X \perp\!\!\!\perp Y|\emptyset$

Properties

For conditional independence:

- Symmetry: $X \perp\!\!\!\perp Y|Z \Rightarrow Y \perp\!\!\!\perp X|Z$
- Decomposition: $X \perp\!\!\!\perp Y, W|Z \Rightarrow X \perp\!\!\!\perp Y|Z$ and $X \perp\!\!\!\perp W|Z$
- Weak union: $X \perp\!\!\!\perp Y, W|Z \Rightarrow X \perp\!\!\!\perp Y|Z, W$
- Contraction: $X \perp\!\!\!\perp W|Z, Y$ and $X \perp\!\!\!\perp Y|Z \Rightarrow X \perp\!\!\!\perp Y, W|Z$
- Intersection:$X \perp\!\!\!\perp Y|W, Z$ and $X \perp\!\!\!\perp W|Y, Z \Rightarrow X \perp\!\!\!\perp Y, W|Z$

For independence: let $Z = \emptyset$ *e.g.*
$X \perp\!\!\!\perp Y \Rightarrow Y \perp\!\!\!\perp X$
$X \perp\!\!\!\perp Y, W \Rightarrow X \perp\!\!\!\perp Y$ and $X \perp\!\!\!\perp W$
...

Marginal and MAP Queries

Given joint distribution $P(Y, E)$, where

- Y, query random variable(s), unknown
- E, evidence random variable(s), observed *i.e.* E = e.

Two types of queries:

- Marginal queries (a.k.a. probability queries)
  task is to compute $P(Y|E = e)$
- MAP queries (a.k.a. most probable explanation)
  task is to find $y^* = \mathrm{argmax}_{y \in Val(Y)} P(Y|E = e)$

That's all

Thanks!