Probability theory

LING 570
Fei Xia

Basic concepts

- Possible outcomes, sample space, event, event space
- Random variable and random vector
- Conditional probability, joint probability, marginal probability

Some terms

- Outcome: a possible result of an experiment
- Sample space: the set of all possible outcomes
- Event: a set of outcomes of an experiment; a subset of the sample space
- Event space: the set of all possible events
- Ex: toss a coin three times and look at head/tail sequence

Random variable

- The outcome of an experiment need not be a number.
- We often want to represent outcomes as numbers.
- A random variable X is a function from the sample space to real numbers: $\Omega \rightarrow R$.
  - Ex: the number of heads with three tosses: X(HHT)=2, X(HTH)=2, X(HTT)=1, …

Two types of random variables

- Discrete: X takes on only a countable number of possible values.
  - Ex: Toss a coin three times. X is the number of heads that are noted.

- Continuous: X takes on an uncountable number of possible values.
  - Ex: X is the speed of a car (e.g., 56.5 mph)

Common distributions

- Discrete random variables:
  - Uniform: all outcomes have the same prob
  - Bernoulli: P(X=1), two possible outcomes
  - Binomial: P(X=k), a trial has two possible outcomes, n trials
  - Multinomial: P(X1=k1, X2=k2, …), a trial has k possible outcomes, n trials
  - Poisson: a good approximation of binomial when n is big and p is small.

- Continuous random variables:
  - Continuous uniform distribution
  - Gaussian (aka normal distribution)
  - …

Random vector

- Random vector is a finite-dimensional vector of random variables: $X=[X_1,\ldots,X_k]$.

- $P(x) = P(x_1,x_2,\ldots,x_n)=P(X_1=x_1,\ldots., X_n=x_n)$

- Ex: $P(w_1, \ldots, w_n, t_1, \ldots, t_n)$

Notation

- X, Y: random variables or random vectors.
- x, y: some values

- $P(X=x)$ is often written as $P(x)$
- $P(X=x \mid Y=y)$ is written as $P(x \mid y)$

Three types of probability

- Joint prob P(x,y): the prob of X=x and Y=y happening together
- Conditional prob P(x | y): the prob of X=x given a specific value of Y=y
- Marginal prob P(x): the prob of X=x for all possible values of Y.

Chain rule: calc joint prob from marginal and conditional prob

$$P(A,B) = P(A) * P(B \mid A) = P(B) * P(A \mid B)$$

$$P(A_1,\ldots,A_n) = \prod_{i>=1} P(A_i \mid A_1,\ldots A_{i-1})$$

Calculating marginal probability from joint probability

$$P(A) = \sum_{B} P(A, B)$$

$$P(A_1) = \sum_{A_2, \ldots, A_n} P(A_1, \ldots, A_n)$$

Bayes' rule

$$P(B \mid A) = \frac{P(A,B)}{P(A)} = \frac{P(A \mid B)P(B)}{P(A)}$$

$$y^* = \arg\max_y P(y \mid x)$$

$$= \arg\max_y \frac{P(x \mid y)P(y)}{P(x)}$$

$$= \arg\max_y P(x \mid y)P(y)$$

Independent random variables

- Two random variables X and Y are independent iff the value of X has no influence on the value of Y and vice versa.
- $P(X,Y) = P(X)\,P(Y)$
- $P(Y \mid X) = P(Y)$
- $P(X \mid Y) = P(X)$

Conditional independence

Once we know C, the value of A does not affect the value of B and vice versa.

- $P(A,B \mid C) = P(A \mid C)\, P(B \mid C)$
- $P(A \mid B,C) = P(A \mid C)$
- $P(B \mid A, C) = P(B \mid C)$

Independence and conditional independence

- If A and B are independent, are they conditional independent?

- Example:
  - Burglar, Earthquake
  - Alarm

Independence assumption

$$P(A_1,\ldots,A_n) = \prod_{i>=1} P(A_i \mid A_1,\ldots A_{i-1})$$

$$\approx \prod_{i>=1} P(A_i \mid A_{i-1})$$

An example

- $P(w_1\ w_2\ \ldots\ w_n)$
  $= P(w_1)\ P(w_2 \mid w_1)\ P(w_3 \mid w_1\ w_2) * \ldots$
  $* P(w_n \mid w_1\ \ldots,\ w_{n-1})$
  $\approx P(w_1)\ P(w_2 \mid w_1)\ \ldots.\ P(w_n \mid w_{n-1})$

- Why do we make independence assumption which we know are not true?

Summary of elementary probability theory

- Basic concepts: sample space, event space, random variable, random vector
- Joint / conditional / marginal probability
- Independence and conditional independence
- Four common tricks:
  - Chain rule
  - Calculating marginal probability from joint probability
  - Bayes' rule
  - Independence assumption