STT 465

Lecture 2:

Belief, probability, independence, conditional independence, exchangeability

G. de los Campos

Beliefs & Probability

⇒ Beliefs and probability (read the chapter in the book)

⇒ Review of probability

(1) Events $\{A_i\}$ & Sample space $S=\{A_1,\ldots\}$

(2) Probability (map from events to numbers in the [0,1] interval that follows a few rules).

(3) $P(A_i)\geq 0$

(4) $P(S)=1$

(5) Probability of the union: $P(A_i U A_j)=P(A_i)+P(A_j)-P(A_i \cap A_j)$ [U=OR, $\cap$=&]

⇒ Marginal, conditional and joint probabilities

(6) Marginal probability: $P(A)$

(7) Joint probability: $P(A\&B)=P(A,B)$ [we will commonly use ',' for joint]

(8) Conditional probability: $p(A|B)$

(9) Factorization: $P(A,B)=P(A)\,P(B|A)=P(B)P(A|B)$

(10) Rule of marginal probability: $P(A\ )=P(A)P(A|B) + P(\text{Not } B)P(B|\text{Not } A)$

⇒ Bayes rule: $p(A|B)=P(B|A)P(A)/P(B)$

Examples

⇒ Review income/education example (p 16)

⇒ Example: genetics of disease

Genotype	Status
AA	H
AB | BA	H
BB	D

$p(A)=0.8$

⇒ Determine

(1) The (marginal) probability of each genotype under random mating, $p(G)$
(2) The probability of disease of a randomly sampled individual, $P(D)$
(3) The conditional probability of disease given genotypes, $p(D|G)$
(4) The joint probability of disease and genotypes, $p(D,G)$

Bayesian Learning

$\Rightarrow$ Compute the probability that a parent is a carrier given that the first offspring is healthy $p(P=AB|O_1=H)$

$\Rightarrow$ Ho would that probability be updated if you know that the second offspring is also healthy?

$\Rightarrow$ What about if you know that the third offspring developed the disease?

Conditional Independence

$\Rightarrow$ X,Y are independent if p(X,Y)=p(X)p(Y)

$\Rightarrow$ X,Y are said <u>conditionally independent</u> if, given a third variable (Z), X and Y are independent, that is:

<u>Conditional Independence:</u> $p(X,Y|Z)=P(X|Z)*P(Y|Z)$

$\Rightarrow$ Note: conditional independence does not imply marginal independence

$\Rightarrow$ HW1-Q1: Consider the disease model of slide 3, with p(A)=0.8, for this model:

(1.1) Derive the conditional probability of disease given the genotype of one of the parents, that is $p(S_i|G_0)$.

(1.2) Derive and report the joint probability of the disease status of two half sibs (i.e., two individuals that share one parent) given the genotype of the known parent, that is $p(S_1, S_2|G_0)$.

(1.3) Derive and report the joint probability of the disease status of two half sibs ,that is $p(S_1, S_2)$.

(1.4) Are S_1 and S_2 independent?

(1.5) Derive and report $p(S_2|S_1)$.

<u>Notation:</u> G_0 denotes the genotype of the known parent and can take values {AA,AB,BB}.
S_i denotes the health status of the ith progeny (i=1,2), it can take values {H,D}.

Remarks

$\Longrightarrow$ The joint probability contains all the information needed to arrive at the marginal s and conditionals

$\Longrightarrow$ We can arrive at the joint distribution from the marginal only with knowledge of the conditional distributions (under independence this is trivial).

Exchangeability & Conditional Independence

⇒ Bruno de Finetti established an important relationship between exchangeable sequences of RVs and Conditional independence (de Finetti's theorem)

⇒ Exchangeable Sequence of Random Variables

- Sequence of Random Variables $Y_1, Y_2, \ldots, Y_n$
- Joint Distribution $p(Y_1, Y_2, \ldots, Y_n)$
- Permutation: $\pi=\{\pi_1, \pi_2, \ldots, \pi_n\}$

⇒ A sequence of RV is said to be exchangeable if

$$p(Y_{\pi 1}, Y_{\pi 2}, \ldots, Y_{\pi N}) \text{ is the same for any permutation } \pi$$

⇒ Note: sequence of IID RVs are exchangeable (discuss), but the converse is not TRUE

⇒ Importantly, independence is not required for exchangeability (discuss MVN case)

de Finetti's Theorem

$\Rightarrow$ Bruno de Finetti established an important relationship between exchangeable sequences of RVs and Conditional independence (de Finetti's theorem)

$\Rightarrow$ If Y_1, Y_2 ,...., Y_n is an exchangeable sequence of random variables then, for some parameter (θ), some prior density, $p(\theta)$ and some conditional density $p(Y_i|\theta)$, the joint distribution of the sequence can be expressed as:

$$p(Y_1, Y_2, \ldots, Y_n) = \int \left\{ \prod_{i=1}^{n} p(Y_i \mid \theta) \right\} p(\theta) d\theta$$

$\Rightarrow$ Note:

$$\text{[Allways]: } p(Y_1, Y_2, \ldots, Y_n) = \int p(Y_1, Y_2, \ldots, Y_n, \theta) d\theta = \int p(Y_1, Y_2, \ldots, Y_n \mid \theta) p(\theta) d\theta$$

$\Rightarrow$ De Finetti's theorem tells us that if the sequence is exchangeable, we can assume conditional IID

$$p(Y_1, Y_2, \ldots, Y_n \mid \theta) = \prod_{i=1}^{n} p(Y_i \mid \theta)$$