Lecture 11: Computing probabilities for random variables

11.1: Three discrete examples

(i) In one experiment, Mendel *selfed* his pea plants: that is, the parents were $RR \times RR$, or $RW \times RW$, or $WW \times WW$. In order to tell whether the parents were $RW \times RW$ he grew up 10 offspring, and if all were red he assumed the parents were $RR \times RR$. Recall that each offspring of an $RW \times RW$ mating is white with probability 1/4. For each $RW \times RW$ mating, what is the probability Mendel mis-called it as $RR \times RR$?

(ii) In the formation of chromosomes that are copied from the parent to go to the sperm or egg cell (or pollen or seed), a process called *crossing over* occurs. For a segment of chromosome of length L the number of crossovers has a Poisson distribution with parameter L.
(a) For a chromosome of length $L = 2$, what is the probability of no crossovers?
(b) What is the probability of 2 or more crossovers?
(For those interested, a length $L = 1$ corresponds to about 10^8 base-pairs (bp) of DNA.)

(iii) Bacteria have just one parent, and genetically they are close copies of that parent. If we take 2 bacteria from a current pool, we can (in theory) trace back their parents, generation after generation until we find their most recent common ancestor. Suppose that, independently at each generation, two bacteria have the same parent with probability $p = 0.02$.
(a) What is the probability that two randomly chosen bacteria have their most recent common ancestor exactly 3 generations ago?
(b) What is the probability the most recent common ancestor is at least 50 generations ago?

11.2 Four examples of continuous distributions

(i) Assume we have an ideal random number generator, that can generate a number uniformly distributed between 0 and 1. We generate two independent such random numbers.
(a) What is the probability the second number is bigger than the first?
(b) What is the probability the first number is between 0.3 and 0.7 (event A)?
(c) What is the probability the second number is either between 0.1 and 0.2, or between 0.7 and 0.8 (event B)?
(d) What is the probability that both events A and B happen?

(ii) The length of chromosome before a crossover event happens is exponential with parameter $\lambda = 1$.
(a) What is the probability that no crossover will occur in length of chromosome $L = 2$?
(b) What is the probability that the first crossover occurs before length $L = 0.5$?
(Compare 11.1 (ii) (a) and 11.2 (ii) (a)).

(iii) The length of time before two bacteria have a common ancestor is exponential with rate parameter $\lambda = 0.02$ per unit time.
(a) What is the probability that two randomly sampled bacteria have a common ancestor more recent than 3 time units ago?
(b) What is the probability that the most recent common ancestor is at least 50 time units ago?
(Compare 11.1 (iii) (b) and 11.2 (iii) (b).)

Lecture 12: More Conditional probability Ross 3.5

12.1 Conditional probability is a probability

1. $P(D \mid E) = P(D \cap E)/P(E) \geq 0$. (we assume $P(E) > 0$.)
2. $P(\Omega \mid E) = P(\Omega \cap E)/P(E) = P(E)/P(E) = 1$.
3. Note $(\cup_i D_i) \cap E = \cup_i (D_i \cap E)$. So, for disjoint D_i,

$$P(\cup_i D_i \mid E) = P(\cup_i (D_i \cap E))/P(E) = \sum_i P(D_i \cap E)/P(E) = \sum_i P(D_i \mid E).$$

So conditional probabilities satisfy all the probability laws. For example,

$$P((C \cup D) \mid E) = P(C|E) + P(D|E) - P(C \cap D \mid E)$$
$$D_1 \subset D_2 \Rightarrow P(D_1 \mid E) \leq P(D_2 \mid E)$$

12.2 Updating information

(i) Bayes' Theorem (again: see lecture notes 5.3).

Assume $P(D)$ and $P(H)$ are both > 0. Then, by definition,

$$P(D \mid H)\ P(H) = P(D \cap H) = P(H \mid D)\ P(D) \quad \text{or} \quad P(H \mid D) = P(D \mid H)\ P(H)\ /\ P(D)$$

Note also, from the law of total probability

$$P(D) = P(D \cap H) + P(D \cap H^c) = P(D|H)P(H) + P(D \mid H^c)(1 - P(H))$$

(ii) Mutually exclusive and exhaustive events H_i ("hypotheses", or "states of nature")

Suppose H_i has probability $P(H_i)$ and $\sum_{i=1}^k P(H_i) = 1$.

$$P(H_i \mid D) = P(D \mid H_i)P(H_i)/\left(\sum_{j=1}^k P(D \mid H_j)P(H_j)\right)$$

12.3 Updating information sequentially

(i) The probability of new data (Ross Ch 3, simpler version of example 5e)

Suppose again that D and E are independent given each H_i: $P(E \mid D) = \sum_{i=1}^k P(E \mid H_i)P(H_i \mid D)$.

Example: two coins C_1 and C_2, with probability head 1/4 and 3/4. Choose one coin randomly and toss it twice. What is the probability the second toss is heads given the first is heads?

Solution 1: $P(\text{2 nd. head} \mid \text{first head}) = P(\text{both heads})/P(\text{1 st head}) =$

$$((1/4) \times (1/4) \times (1/2) + (3/4) \times (3/4) \times (1/2))/((1/4) \times (1/2) + (3/4) \times (1/2)) = 5/8.$$

Solution 2: After first head, $P^*(C_1) = P(C_1 \mid \text{head}) = (1/4) \times (1/2)/((1/4) \times (1/2) + (3/4) \times (1/2)) = 1/4$.

So $P^*(C_2) = P(C_2 \mid \text{head}) = (1 - (1/4)) = 3/4$.

$P(\text{heads again}) = P^*(C_1)P(\text{head} \mid C_1) + P^*(C_2)P(\text{head} \mid C_2) = (1/4) \times (1/4) + (3/4) \times (3/4) = 5/8$.

(ii) The general case (Ross, Ch 3; example 5f)

Suppose now we have two data events D and E:

$$P(H_i \mid D \cap E) = P(D \cap E \mid H_i)P(H_i)/\left(\sum_{j=1}^k P(D \cap E \mid H_j)P(H_j)\right)$$

Also, provided D and E are independent given each H_i, $i = 1, ..., k$:

$$P(H_i \mid D \cap E) = P(E \mid H_i)P(H_i \mid D)/\left(\sum_{j=1}^k P(E \mid H_j)P(H_j \mid D)\right)$$

That is, we can first update from $P(H_i)$ to $P(H_i \mid D)$ and then use these probabilities in updating to $P(H_i \mid D \cap E)$. And then so also for the next event, and the next, ...