Lecture 4: More counting examples

4.1 Binomial counts and Stirling's formula

(i) Suppose there are N equiprobable outcomes in Ω.

Suppose event E is true for R of these outcomes. Then $P(E) = R/N$.

(ii) An AB parent and an O parent can have an A child or a B child.

Suppose they have n children: there are 2^n possible sequences of A and B children.

Assume these are equiprobable. (In fact, they are.)

$\binom{n}{k}$ of these sequences have k A children. $P(k$ A children out of $n) = \binom{n}{k}/2^n$.

(iii) $n!$ can be approximated for large n by $\sqrt{2\pi}n^{n+\frac{1}{2}}e^{-n}$. Also $\binom{n}{k}$ is largest when $k \approx n/2$.

Then, for large n,
$$\binom{n}{n/2} = \frac{n!}{(n/2)!(n/2)!} \approx \frac{\sqrt{2\pi}n^{n+\frac{1}{2}}e^{-n}}{\sqrt{2\pi}(n/2)^{(n/2)+\frac{1}{2}}e^{-(n/2)} \times \sqrt{2\pi}(n/2)^{(n/2)+\frac{1}{2}}e^{-(n/2)}}$$
$$= (1/\sqrt{2\pi})2^{n+1}n^{n+\frac{1}{2}-(n/2)-\frac{1}{2}-(n/2)-\frac{1}{2}} = (1/\sqrt{2\pi})(2/\sqrt{n})2^n$$

Or $P((n/2)$ A children out of $n) = \binom{n}{n/2}(\frac{1}{2})^n = 1/\sqrt{2\pi(n/4)}$

This result will have meaning later when we discuss approximating Binomial probabilities by the Normal probability distribution.

4.2 The binomial theorem
$$(x+y)^n = \sum_{k=0}^{n}\binom{n}{k}x^k y^{n-k}$$

Proof by induction is given in Ross, P 8-9. It uses the formula of notes 3.2(v).

Note in each bracket we choose x or y. There are 2^n sequences.

The number of sequences in which there are k choices of x is $\binom{n}{k}$, and each has value $x^k y^{n-k}$.

The case of 4.1 (ii) is a special case when $x = y = \frac{1}{2}$.

4.3 Sampling with and without replacement (Ross 2.5)

(i) Draw 3 cards from 52-card pack. $E = \{$draw at least one face card$\}$. Note E^c (no face card) is easier.

With replacement: P(no face card) $= (40 \times 40 \times 40)/(52 \times 52 \times 52) = 0.455$.

P(at least 1 face card) $= 1 - 0.455 = 0.545$.

Without replacement: P(no face card) $= (40 \times 39 \times 38)/(52 \times 51) \times 50) = 0.447$.

(ii) **The birthday problem:** ignore Feb 29, and assume other days equiprobable.

In k people, E is event that at least 2 share a birthday. E^c is the event that all bithdays are different.

$P(E^c) = 365.364.363.362......(365 - k + 1)$ / $365.365.365.....365 = 365!$ / $(365 - k)! \times (1/365)^k$

$k = 2$, $P(E) = 1/365$; $k = 23$, $P(E) \approx 0.5$; $k = 45$, $P(E) = 0.94$.

(iii) **Hypergeometric probabilities:** N fish in a pond; n are red, $N - n$ are blue.

Sample k fish without replacement. What is the probability x are red?

Total number of outcomes is $\binom{N}{k}$. Number of ways of choosing the x red fish is $\binom{n}{x}$, and the $(k - x)$ blue

fish is $\binom{N-n}{k-x}$. Probability is $\binom{n}{x} \times \binom{N-n}{k-x} / \binom{N}{k}$

Lecture 5: Conditional probability and independence Ross 3.2,3.3.

5.1 Conditional probability (Ross 3.2)

(i) Idea: revise probabilities in light of restrictions or partial information.

(ii) Definition: probability of D given E is $P(D \mid E) = P(D \cap E)/P(E)$, provided $P(E) > 0$.

(iii) Example: Our AB and O parents have three children.

$\Omega = \{AAA, AAB, ABA, ABB, BAA, BAB, BBA, BBB\}$ and these are equiprobable (1/8 each).

Let D be at least 2 A children: $P(D) = 4/8 = 1/2$. Note also $P(\text{first child is } A) = 4/8 = 1/2$.

$P(D \mid \text{first child is } A) = P(D \cap \text{first child is } A)/P(\text{first child is } A) = (3/8)/(4/8) = 3/4$.

5.2 The multiplication rule (Ross 3.2) If $P(E_1 \cap E_2 \cap ... \cap E_n) > 0$,

$$P(E_1 \cap E_2 \cap ... \cap E_n) = P(E_1).P(E_2|E_1).P(E_3 \mid E_1 \cap E_2)....P(E_n \mid E_1 \cap E_2 \cap ...... \cap E_n)$$

To prove, just write it out:

$$P(E_1).P(E_2|E_1).P(E_3 \mid E_1 \cap E_2)....P(E_n \mid E_1 \cap E_2 \cap ...... \cap E_n) =$$
$$P(E_1)\frac{P(E_2 \cap E_1)}{P(E_1)}\frac{P(E_3 \cap E_2 \cap E_1)}{P(E_2 \cap E_1)}... \qquad ...\frac{P(E_n \cap ... \cap E_3 \cap E_2 \cap E_1)}{P(E_{n-1} \cap .. \cap E_2 \cap E_1)}$$

This looks messy, but in fact we have already used it in examples of sampling without replacement.

Example: No face card in 3 cards, when sampling without replacement. $(40/52).(39/51).(38/50)$.

5.3 Bayes' formula

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

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

Note also, from the law of total probability, we could compute $P(E)$ as

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

Example: $P(\text{first child is } A \mid D \equiv \text{"at least } 2A \text{ children"})$
$$= P(D \mid \text{first is } A)P(\text{first is } A)/P(D) = (3/4) \times (1/2) \,/\, (1/2) = 3/4.$$

For better examples see Lecture 6: $P(D|A) = P(A|D)$ **ONLY BECAUSE** $P(D) = P(A)$.

5.4 Independent events (Ross 3.4)

(i) Definition: E and F are independent if $P(E \cap F) = P(E) \times P(F)$.

(ii) Intepretation: Knowing F happens does not affect the probability of E: $P(E|F) = P(E)$.

(iii) If E and F are independent, $P(E \cup F) = P(E) + P(F) - P(E).P(F)$ so

$P(E^c \cap F^c) = 1 - P(E \cup F) = 1 - P(E) - P(F) + P(E).P(F) = (1 - P(E)).(1 - P(F)) = P(E^c).P(F^c)$.

That is E^c and F^c are independent. (So are E^c and F, and E and F^c).

Example: $P(A) = 0.36$, $P(B) = 0.2$, $P(AB) = 0.08, P(O) = 0.36$.

$P(\text{have antigen } A) = P(A) + P(AB) = 0.44$, $P(\text{have antigen } B) = P(B) + P(AB) = 0.28$.

$0.44 \times 0.28 = 0.1232 \neq 0.08 = P(AB)$: Having antigen A is NOT independent of having antigen B.

$P(\text{have antigen } A \mid \text{have antigen } B) = P(AB)/P(\text{have antigen } B) = 0.08/0.28 = 0.286$
$$< 0.44 = P(\text{have antigen } A)$$

Knowing a person has the B antigen decreases the probability they have the A antigen.

Lecture 6: Probability Examples

6.1 More conditional probabilities

(i) Mendel discovered that every pea plant has two *factors* or *alleles* for flower color: R (red) and W (white). Each plant has RR or RW and is red, or WW and is white.

We cross two plants which we know are RW.

The offspring plants are *independent* and each is RR, RW or WW with probabilities 1/4, 1/2, 1/4.

What is the probability an offspring is RR, given it is red?

Answer: $P(\text{red}) = 3/4$, $P(RR \cap \text{red}) = P(RR) = 1/4$, so $P(RR \mid \text{red}) = (1/4)/(3/4) = 1/3$.

(ii) **Testing for a rare disease** (Ross Ch 3, example 3d)

Suppose we have a quite effective test, so $P(+ \mid \text{disease}) = 0.99$, and test is quite accurate, so $P(+ \mid \text{no disease}) = 0.02$). Now suppose the frequency of the disease is 0.001.

$$
\begin{aligned}
P(+ \text{ test result}) &= P(+ \mid \text{disease})P(\text{disease}) + P(+ \mid \text{no disease})P(\text{no disease}) \\
&= 0.99 \times 0.001 + 0.02 \times (1 - 0.001) = 0.00099 + 0.0198 = 0.02097 \\
P(\text{disease} \mid +) &= P(+ \mid \text{disease})P(\text{disease})/P(+) = 0.99 \times 0.001/0.02097 = 0.047
\end{aligned}
$$

Less than 5% of people testing positive actually have the disease!

• The probability of spots given measles is large: the probability of measles given spots is small.

• The probability of 4 legs given elephant close to 1: the probability of elephant given 4 legs is close to 0.

6.2 Updating with more information

(i) Allele frequencies in a population of pea plants.

Suppose the R allele has frequency 0.3, and W 0.7, and the types of the two alleles in an individual are *independent*. What is the probability a red-flowered pea-plant is RR?

Solution: By independence, $P(RR) = 0.3 \times 0.3 = 0.09$, $P(WW) = 0.7 \times 0.7 = 0.49$, $P(RW) = 1 - 0.09 - 0.49 = 0.42$.

Overall $P(\text{red}) = P(RR) + P(RW) = 0.09 + 0.42 = 0.51$,

so then $P(RR \mid \text{red}) = P(RR \cap \text{red})/P(\text{red}) = 0.09/0.51 = 3/17$. $P(RW \mid \text{red}) = 1 - (3/17) = 14/17$.

(ii) The red pea plant is crossed to a white one (WW). The first offspring has red flowers. What now is the probability the red parent plant is RR?

Solution: $P(\text{red offspring} \mid RR \text{ parent}) = 1$, $P(\text{red offspring} \mid RW \text{ parent}) = (1/2)$.

$P(\text{red offspring}) = P(\text{red offspring} \mid RR)P(RR) + P(\text{red offspring} \mid RW)P(RW)$
$$= 1 \times (3/17) + (1/2) \times (14/17) = 10/17.$$

$P(\text{red parent is } RR \mid \text{red offspring}) = P(\text{red offspring} \mid \text{red parent is } RR) \times P(\text{red parent is } RR)/P(\text{red offspring})$
$$= 1 \times (3/17)/(10/17) = 0.3.$$

6.3 Independent events

(i) Back to the two RW parent pea plants of 6.1 (i):

Each offspring is red with probability 3/4, and white with probability 1/4, and the colors of offspring pea plants are *independent*. We grow up 10 offspring.

What is the probability all 10 are red: answer $(3/4)^{10} = 0.0563$

What is the probability of at least one red: answer $1 - P(\text{all white}) = 1 - (1/4)^{10} = 0.9999$.