Dependence & Independence

Statistics 110

Summer 2006

Dependence & Independence

Biased coin example revisited:

Flip a coin until you see a tail and let X be the flip number when this occurs ($\Omega = \{1, 2, 3, \ldots\}$). On any particular flip, $P[H] = \frac{2}{3}$ and $P[T] = \frac{1}{3}$. The model proposed had $P[X = i]$ satisfying

$$P[X = i] = \frac{1}{3}\left(\frac{2}{3}\right)^{i-1}; i = 1, 2, \ldots$$

Where did this come?

For $X = i$, the flips must be

$$\underbrace{H\ H\ H \ldots H}_{i-1 \text{ of them}} T$$

So

$$P[X = i] = P[\underbrace{H\ H\ H \ldots H}_{i-1 \text{ of them}} T]$$

In this, order matters. Let F_j be the result on flip j. Then by the multiplication rule,

$$\begin{aligned}
&P[X = i] = \\
&\qquad P[F_1 = H] \times P[F_2 = H | F_1 = H] \times P[F_3 = H | F_1 = H, F_2 = H] \ldots \\
&\qquad\quad \times P[F_{i-1} = H | F_1 = H, \ldots, F_{i-2} = H] \\
&\qquad\quad \times P[F_i = T | F_1 = H, \ldots, F_{i-1} = H]
\end{aligned}$$

Under the assumption made,

$$P[F_j = H | \text{Earlier flips}] = \frac{2}{3}$$

What happens on flip j isn't influenced by any of the earlier flips. So in above formula $i-1$ of the terms are $\frac{2}{3}$ and the last term is $\frac{1}{3}$, thus giving

$$P[X=i] = \frac{1}{3}\left(\frac{2}{3}\right)^{i-1}; i = 1, 2, \ldots$$

This coin flipping example is equivalent to sampling from the population with 20 Heads and 10 Tails with replacement. What happens if we sample without replacement? What does $P[X=i]$ look like in this case?

The relationship

$$\begin{aligned} P[X=i] = P[F_1 = H] \times P[F_2 = H|F_1 = H] \times \ldots \\ \times P[F_i = T|F_1 = H, \ldots, F_{i-1} = H] \end{aligned}$$

is still valid.

However $P[F_j = H|\text{Earlier flips}]$ isn't always $\frac{2}{3}$. For example

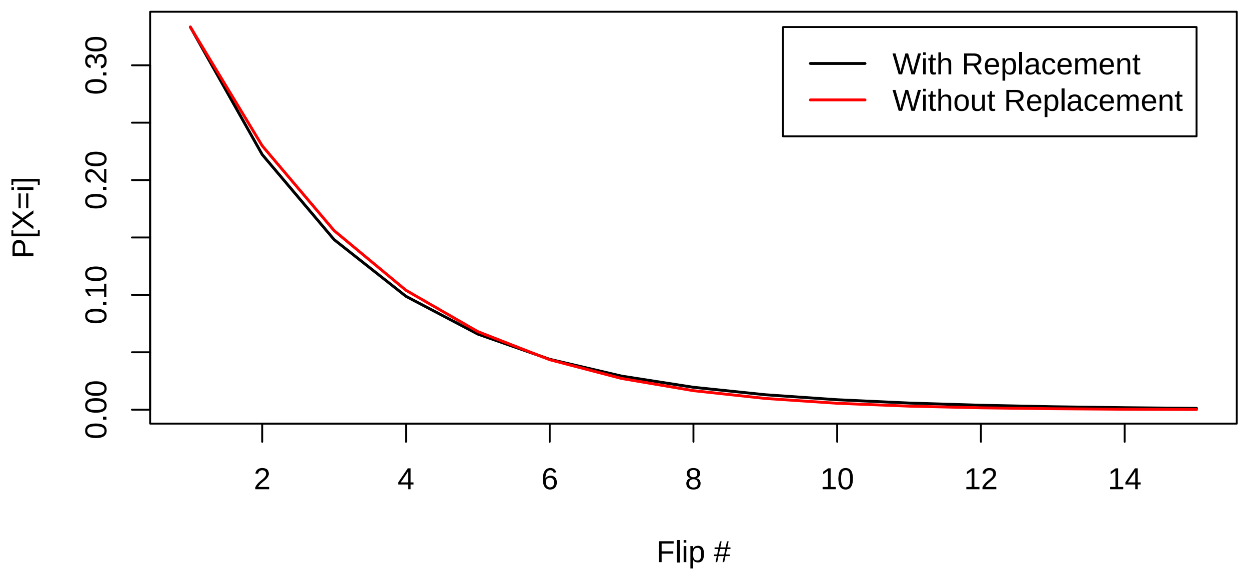
$$P[F_2 = H|F_1 = H] = \frac{19}{29} < \frac{2}{3}$$
$$P[F_2 = H|F_1 = T] = \frac{20}{29} > \frac{2}{3}$$

Based on this,

$$P[X = i] = \begin{cases} \frac{1}{3} & i = 1 \\ \frac{20 \times 19 \times \ldots \times (20 - i + 2)}{30 \times 29 \times \ldots \times (30 - i + 2)} \times \frac{10}{30 - i + 1} & i > 1 \end{cases}$$

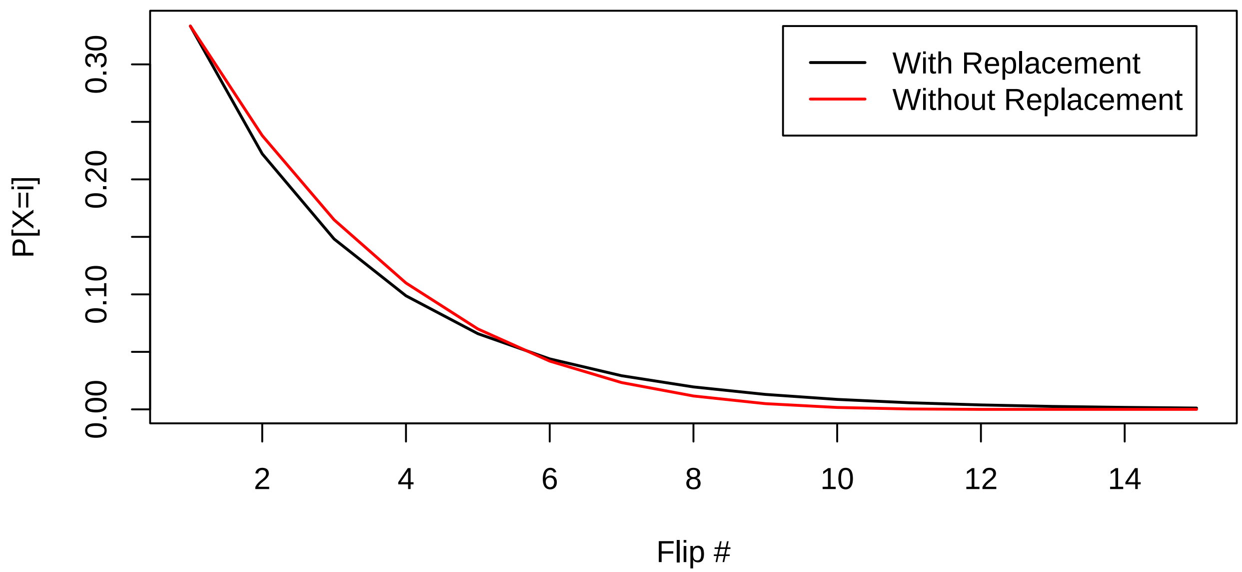
$P[X = i]$ under both sampling schemes (20 Heads - 10 Tails)

i	With Replacement	Without Replacement
1	0.3333	0.3333
2	0.2222	0.2299
3	0.1482	0.1560
4	0.0988	0.1040



$P[X = i]$ under both sampling schemes (10 Heads - 5 Tails)

i	With Replacement	Without Replacement
1	0.3333	0.3333
2	0.2222	0.2381
3	0.1482	0.1648
4	0.0988	0.1099



The draws in the with replacement sampling scheme are said to be **Independent** whereas the draws in the without replacement scheme are **Dependent**.

In general, the idea behind independence is that two events are independent if knowledge about one event occurring gives us no information about whether the other event occurred.

In the with replacement scheme, knowing that a head occurs on the first draw doesn't change anything for the second draw since there are still 20 Heads and 10 Tails in the bag for the second draw. However for the without replacement scheme, the results of the first draw changes the probabilities for the second draw.

Definition:

Two events A and B are said to be **Independent** if

$$P[A \cap B] = P[A]P[B]$$

Events that are not independent are often said to be **dependent**.

Back to sampling example

- With Replacement

$$P[F_1 = H, F_2 = H] = \frac{4}{9} = \frac{2}{3} \times \frac{2}{3} = P[F_1 = H]P[F_2 = H]$$

so the events $\{F_1 = H\}$ and $\{F_2 = H\}$ are independent.

- Without Replacement

$$P[F_1 = H, F_2 = H] = \frac{20 \times 19}{30 \times 29} = 0.4368$$

$$P[F_1 = H] = \frac{20}{30} = \frac{2}{3}$$

$$P[F_2 = H] = \frac{20}{30} \times \frac{19}{29} + \frac{10}{30} \times \frac{20}{29} = \frac{2}{3}$$

$$P[F_1 = H]P[F_2 = H] = 0.4444 \neq P[F_1 = H, F_2 = H]$$

so the events $\{F_1 = H\}$ and $\{F_2 = H\}$ are not independent under without replacement sampling.

We can also think of independence in terms of conditional probability. The following lemma justifies it

Lemma. *Suppose* $0 < P[A] < 1$. *Then the following three conditions are equivalent:*

- $P[A \cap B] = P[A]P[B]$
- $P[B|A] = P[B]$
- $P[B|A] = P[B|A^c]$

When dealing with independence, there is much symmetry. For example if $0 < P[B] < 1$, then $P[A|B] = P[A]$. Also if A and B are independent events, so are A^c and B, A and B^c, and A^c and B^c.

Many of the other examples I've discussed the events of interest have been dependent. For example, in the ELISA example,

$$P[\text{HIV}|+\text{test}] = 0.124$$
$$P[\text{HIV}|-\text{test}] = 0.00022$$
$$P[\text{HIV}] = 0.01$$

In the juror example,

$$P[\text{Unbiased}|\text{Excused with cause}] = 0$$
$$P[\text{Unbiased}|\text{Not excused with cause}] = 0.7117$$
$$P[\text{Unbiased}] = 0.5$$

Questions:

1. Are A and A^c independent?

2. Assume that A and B are disjoint events. Are they independent?

Independence with more than two events

How do we want to define independence in this case?

Examples:

1. Suppose two fair dice (one red and one blue) are rolled and the following events are defined

   - A: a 1 or 2 on the red die
   - B: a 3, 4, or 5 on the blue die
   - C: total is 4, 11, or 12.

Its possible to show that

$$P[A] = \tfrac{1}{3} \qquad P[B] = \tfrac{1}{2} \qquad P[C] = \tfrac{1}{6}$$

$$P[A \cap B] = \tfrac{1}{6} \qquad P[A \cap C] = \tfrac{1}{18} \quad P[B \cap C] = \tfrac{1}{18}$$

$$P[A \cap B \cap C] = \tfrac{1}{36}$$

While

$$P[A \cap B] = P[A]P[B]$$
$$P[A \cap C] = P[A]P[C]$$
$$P[A \cap B \cap C] = P[A]P[B]P[C]$$

for the last pairwise case,

$$P[B \cap C] \neq P[B]P[C]$$

2. Roll the same two dice again and define the following events

- D: red die is odd
- E: blue die is odd
- F: total is odd

$$
\begin{array}{ccc}
P[D] = \frac{1}{2} & P[E] = \frac{1}{2} & P[F] = \frac{1}{2} \\
P[D \cap E] = \frac{1}{4} & P[D \cap F] = \frac{1}{4} & P[E \cap F] = \frac{1}{4} \\
P[D \cap E \cap F] = 0 & &
\end{array}
$$

So

$$P[D \cap E] = P[D]P[E]$$
$$P[D \cap F] = P[D]P[F]$$
$$P[E \cap F] = P[E]P[F]$$

but for the three-way case,

$$P[A \cap B \cap C] \neq P[D]P[E]P[F]$$

So for neither of these two situation do we want to consider all the events independent.

The case where every pair of events is independent (like the 2nd example) is known as **Pairwise Independence**.

So you can have different possible independence patterns with more than two events.

Definition:

We define a collection of events $A_1, A_2, \ldots, A_n$ to be **Mutually Independent** if for every subcollection of events $A_{i_1}, A_{i_2}, \ldots, A_{i_m}$,

$$P[A_{i_1} \cap \ldots \cap A_{i_m}] = P[A_{i_1}] \times \cdots \times P[A_{i_m}]$$

When building models, independence is almost always an assumption. In the coin flipping example, its probably a reasonable assumption. In other cases it can be a poor assumption.

Example: People vs Collins

In a 1964 case in a California Court, a conviction for purse snatching in Los Angeles was based on the use of the multiplication rule assuming a number of events were all independent. The victim described her assailant as a young, blond female with a pony tail. The suspect fled on foot but was seen shortly thereafter getting into a yellow car drive by a black man who had a mustache and a beard.

The police investigation turned up a suspect who was blond with a pony tail and associated with a black man who drove a yellow car and had a beard and mustache.

As there was no tangible evidence or reliable witnesses, the prosecutor built his case on how unlikely it would be find pair satisfying the 6 conditions observed of the assailant. The following probabilities were assigned:

$$P[\text{Yellow car}] = 0.1 \qquad P[\text{Man with mustache}] = 0.25$$
$$P[\text{Woman with pony tail}] = 0.1 \qquad P[\text{Woman with blond hair}] = 0.33$$
$$P[\text{Black man with beard}] = 0.1 \qquad P[\text{Interracial couple in car}] = 0.001$$

The prosecutor multiplied these 6 numbers together and claimed that finding another couple like this had a probability of 1 in 12,000,000. Since this probability is so small, the prosecutor claimed it must be case that the defendant is guilty. The jury agreed and convicted the defendant of second-degree robbery.

The California Supreme Court disagreed, and overturned the conviction. They claimed that the probability argument was incorrect and misleading.

Some of these events are probably highly dependent, such as black man with a beard and a man with a mustache.

Another issue addressed in the appeal decision, was there no evidence to support the assumed marginal probability values. However, even if these values were reasonable, it is quite likely that the chance of finding a couple satisfying these 6 conditions is higher than 1 in 12,000,000 (my guess, not from the court's decision).

Similar issues were brought up in the early use of DNA fingerprint data for forensic purposes (such as was used in the OJ trial). The questions included

- What are the correct probabilities for each marker? Do you need to worry about racial background? (Do you need to use conditional probabilities?)

- Is Hardy-Weinburg reasonable? (Is marker inherited from the mother independent of the marker inherited from the father?)

- Is linkage equilibrium reasonable? (Is one marker location independent of another marker location?)

Now however, the issues are effectively moot as lab techniques have eliminated the use of probability calculation to assess the importance of a DNA match.

They are now using so many markers, the only way a second person could match a sample is if they were an identical twin (assuming no lab errors).