Independence – Solutions

STAT-UB.0103 – Statistics for Business Control and Regression Models

The Birthday Problem

1. A class has 70 students. What is the probability that at least two students have the same birthday? Assume that each person in the class was assigned a random birthday between January 1 and December 31.

Solution: Assume that everyone in the class is randomly assigned a birthday, which corresponds to number between 1 and 365 representing the day of the year. It turns out to be much easier to compute the probability using the complement rule, as

$$\text{P(at least 2 people have the same birthday)} = 1 - \text{P(all 70 birthdays are different)}.$$

The next trick is to write the event that all 70 birthdays are different in a redundant way:

$$\begin{aligned}\{\text{all 70 birthdays are different}\} &= \{\text{first 2 are different}\} \cap \{\text{first 3 are different}\} \\ &\cap \{\text{first 4 are different}\} \cap \{\text{first 5 are different}\} \cap \cdots \cap \{\text{first 70 are different}\}.\end{aligned}$$

In class we showed how to use the multiplicative rule repeatedly to get:

$$\begin{aligned}\text{P(all 70 birthdays are different)} &= \text{P(first 2 diff.)} \cdot \text{P(first 3 diff. | first 2 diff.)} \\ &\cdot \text{P(first 4 diff. | first 3 diff.)} \cdot \text{P(first 5 diff. | first 4 diff.)} \cdot \cdots \cdot \text{P(first 70 diff. | first 69 diff.)}.\end{aligned}$$

Using this expression we can compute

$$\text{P(all 70 birthdays are different)} = \frac{364}{365} \cdot \frac{363}{365} \cdot \frac{362}{365} \cdot \cdots \cdot \frac{365-69}{365}.$$

We can do a similar calculation for other class sizes. The following table shows the probabilities of having at least two students with the same birthday for various class sizes:

Class Size	P(all diff.)	P(at least 2 same)
10	88%	12%
20	59%	41%
30	29%	71%
40	11%	89%
50	3%	97%
60	0.6%	99.4%
70	0.08%	99.92%

Independence

2. Suppose you flip two fair coins. Let $A =$"the first coin shows Heads," $B =$ "The second coin shows Heads." Find the probability of getting Heads on both coins, i.e. find $P(A \cap B)$.

Solution: The long way to solve this problem is to write out the elementary outcomes and their probabilities:

Outcome	Probability
HH	$\frac{1}{4}$
HT	$\frac{1}{4}$
TH	$\frac{1}{4}$
TT	$\frac{1}{4}$

Since $A \cap B = \{\text{HH}\}$, it follows that

$$\mathrm{P}(A \cap B) = \tfrac{1}{4}.$$

We can solve this problem much more expediently using the independence of A and B:

$$\begin{aligned}\mathrm{P}(A \cap B) &= \mathrm{P}(A)\,\mathrm{P}(B \mid A) \\ &= \mathrm{P}(A)\,\mathrm{P}(B) \\ &= (\tfrac{1}{2})(\tfrac{1}{2}) \\ &= \tfrac{1}{4}.\end{aligned}$$

3. Suppose you roll three dice. What is the probability of getting exactly one 6?

Solution: Define the following events:

$$\begin{aligned}A &= \text{“6 on the first roll,”} \\ B &= \text{“6 on the second roll,”} \\ C &= \text{“6 on the third roll.”}\end{aligned}$$

Using the shorthand $\bar{A} = A^c$ and $A\bar{B}\bar{C} = A \cap \bar{B} \cap \bar{C}$, the event "exactly one 6" can be written as

$$\text{“exactly one 6”} = A\bar{B}\bar{C} \cup \bar{A}B\bar{C} \cup \bar{A}\bar{B}C.$$

These events are mutually independent, so

$$\mathrm{P}(\text{exactly one 6}) = \mathrm{P}(A\bar{B}\bar{C}) + \mathrm{P}(\bar{A}B\bar{C}) + \mathrm{P}(\bar{A}\bar{B}C).$$

Using the independence of events A, B, and C, we get

$$\begin{aligned}
\mathrm{P}(A\bar{B}\bar{C}) &= \mathrm{P}(A)\,\mathrm{P}(\bar{B})\,\mathrm{P}(\bar{C}) = (\tfrac{1}{6})(\tfrac{5}{6})(\tfrac{5}{6}) \\
\mathrm{P}(\bar{A}B\bar{C}) &= \mathrm{P}(\bar{A})\,\mathrm{P}(B)\,\mathrm{P}(\bar{C}) = (\tfrac{5}{6})(\tfrac{1}{6})(\tfrac{5}{6}) \\
\mathrm{P}(\bar{A}\bar{B}C) &= \mathrm{P}(\bar{A})\,\mathrm{P}(\bar{B})\,\mathrm{P}(C) = (\tfrac{5}{6})(\tfrac{5}{6})(\tfrac{1}{6})
\end{aligned}$$

Note that these three expressions are all equal. Thus,

$$\mathrm{P}(\text{exactly one 6}) = 3 \cdot \tfrac{1}{6} \cdot \tfrac{5}{6} \cdot \tfrac{5}{6} \approx .347.$$

4. Suppose you roll two dice, one red and one green. Let $A =$ "The sum is 7," $B =$ "The red die is a 6." Are events A and B independent?

Solution: Let (r, g) denote the elementary outcome where the red die takes the value r and the green die takes the value g. Events A, B, and their intersection are given by

$$\begin{aligned}
A &= \{(1,6),(2,5),(3,4),(4,3),(5,2),(6,1)\} \\
B &= \{(6,1),(6,2),(6,3),(6,4),(6,5),(6,6)\} \\
A \cap B &= \{(6,1)\}.
\end{aligned}$$

There are 6 possible values for the red die, and 6 possible values for the green die, so there are $6 \cdot 6 = 36$ total elementary outcomes (i.e., $|\Omega| = 36$). Since all elementary outcomes are equally likely, we have

$$\begin{aligned}
\mathrm{P}(A) &= \frac{|A|}{|\Omega|} = \frac{6}{36}, \\
\mathrm{P}(B) &= \frac{|B|}{|\Omega|} = \frac{6}{36}, \\
\mathrm{P}(A \cap B) &= \frac{|A \cap B|}{|\Omega|} = \frac{1}{36}.
\end{aligned}$$

We compute

$$\mathrm{P}(A)\,\mathrm{P}(B) = \left(\frac{6}{36}\right)^2 = \frac{1}{36}.$$

Since $\mathrm{P}(A)\,\mathrm{P}(B) = \mathrm{P}(A \cap B)$, it follows that A and B are independent.

5. Suppose you roll two dice, one red and one green. Let $A =$ "The sum is 8," $B =$ "The red die is a 6." Are events A and B independent?

Solution: We use the same notation as in the solution to problem 4. Events A, B, and their intersection are given by

$$\begin{aligned} A &= \{(2,6),(3,5),(4,4),(5,3),(6,2)\} \\ B &= \{(6,1),(6,2),(6,3),(6,4),(6,5),(6,6)\} \\ A \cap B &= \{(6,2)\}. \end{aligned}$$

The probabilities of these events are

$$\begin{aligned} \mathrm{P}(A) &= \frac{|A|}{|\Omega|} = \frac{5}{36}, \\ \mathrm{P}(B) &= \frac{|B|}{|\Omega|} = \frac{6}{36}, \\ \mathrm{P}(A \cap B) &= \frac{|A \cap B|}{|\Omega|} = \frac{1}{36}. \end{aligned}$$

We compute

$$\mathrm{P}(A)\,\mathrm{P}(B) = \left(\frac{5}{36}\right)\left(\frac{6}{36}\right) = \frac{5}{216}.$$

Since $\mathrm{P}(A)\,\mathrm{P}(B) \neq \mathrm{P}(A \cap B)$, it follows that A and B are *not* independent.

6. Suppose you have a database of 300K reviews from 15K businesses and 70K users. In each of the following scenarios, you randomly sample 2 reviews. Define events A and B as

$$A = \text{the first review is 4 or 5 stars}$$
$$B = \text{the second review is 4 or 5 stars}$$

In which sampling schemes are events A and B independent? Assume that all samples are random and unbiased. Explain your answers.

(a) You sample two distinct reviews from the entire dataset.

> **Solution:** Dependent, but very weakly so. (If the reviews are sampled with replacement, then they are independent.)

(b) You randomly sample one business from the dataset, then sample two distinct reviews of the business.

> **Solution:** Dependent. If the first review is high, then the restaurant is likely good, and so the second review is likely high as well.

(c) You randomly sample one user from the dataset, then sample two distinct reviews written by the user.

> **Solution:** Dependent. The first review tells you about the user, and that in turn tells you about the second review.

(d) You randomly sample two distinct users from the dataset, then sample one review written by each user.

> **Solution:** Dependent, but weakly so. (Think of a where the dataset has three users: one user only leaves 1 star reviews, one user only leaves 3 star reviews, and one user only leaves 5 star reviews)

(e) You randomly sample two distinct businesses from the dataset, then sample one review from each business.

> **Solution:** Dependent, but weakly so. (This is similar to the last example.)