Combinations of Events Intro

Note 14

Product rule: We can find the probability of an intersection of events by enforcing an "ordering" of these events. Here, each successive conditional probability in the product finds the probability of the next event, *conditioned* on all prior events occurring:

$$\mathbb{P}[A_1 \cap A_2 \cap \cdots \cap A_n] = \mathbb{P}[A_1]\,\mathbb{P}[A_2 \mid A_1]\,\mathbb{P}[A_3 \mid A_1 \cap A_2] \cdots \mathbb{P}[A_n \mid A_1 \cap A_2 \cap \cdots \cap A_{n-1}].$$

Note that this is just a generalization of the definition of conditional probability: $\mathbb{P}[A_1 \cap A_2] = \mathbb{P}[A_1]\,\mathbb{P}[A_2 \mid A_1]$

Union Bound: Derived from the principle of inclusion-exclusion, the probability that at least one of the events $A_1, A_2, \ldots, A_n$ occurs is at most the sum of the probabilities of the individual events:

$$\mathbb{P}[A_1 \cup A_2 \cup \cdots \cup A_n] \leq \mathbb{P}[A_1] + \mathbb{P}[A_2] + \cdots + \mathbb{P}[A_n]$$

$$\mathbb{P}[\bigcup_{i=1}^{n} A_i] \leq \sum_{i=1}^{n} \mathbb{P}[A_i]$$

with equality when the A_i's are disjoint.

1 Probability Potpourri

Note 13
Note 14

Provide brief justification for each part.

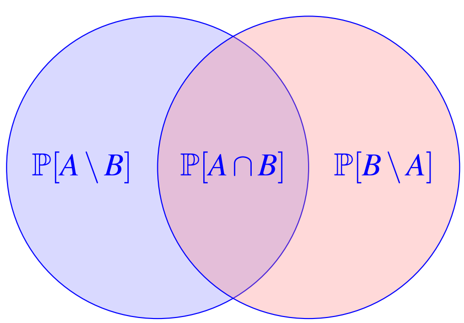
(a) For two events A and B in any probability space, show that $\mathbb{P}[A \setminus B] \geq \mathbb{P}[A] - \mathbb{P}[B]$.

(b) Suppose $\mathbb{P}[D \mid C] = \mathbb{P}[D \mid \overline{C}]$, where $\overline{C}$ is the complement of C. Prove that D is independent of C.

(c) If A and B are disjoint, does that imply they're independent?

Solution:

(a) It can be helpful to first draw out a Venn diagram:



We can see here that $\mathbb{P}[A] = \mathbb{P}[A \cap B] + \mathbb{P}[A \setminus B]$, and that $\mathbb{P}[B] = \mathbb{P}[A \cap B] + \mathbb{P}[B \setminus A]$.

Looking at the RHS, we have

$$\begin{aligned}\mathbb{P}[A] - \mathbb{P}[B] &= (\mathbb{P}[A \cap B] + \mathbb{P}[A \setminus B]) - (\mathbb{P}[A \cap B] + \mathbb{P}[B \setminus A]) \\ &= \mathbb{P}[A \setminus B] - \mathbb{P}[B \setminus A] \\ &\le \mathbb{P}[A \setminus B]\end{aligned}$$

(b) Using the total probability rule, we have

$$\mathbb{P}[D] = \mathbb{P}[D \cap C] + \mathbb{P}[D \cap \overline{C}] = \mathbb{P}[D \mid C] \cdot \mathbb{P}[C] + \mathbb{P}[D \mid \overline{C}] \cdot \mathbb{P}[\overline{C}].$$

But we know that $\mathbb{P}[D \mid C] = \mathbb{P}[D \mid \overline{C}]$, so this simplifies to

$$\mathbb{P}[D] = \mathbb{P}[D \mid C] \cdot (\mathbb{P}[C] + \mathbb{P}[\overline{C}]) = \mathbb{P}[D \mid C] \cdot 1 = \mathbb{P}[D \mid C],$$

which defines independence.

(c) No; if two events are disjoint, we cannot conclude they are independent. Consider a roll of a fair six-sided die. Let A be the event that we roll a 1, and let B be the event that we roll a 2. Certainly A and B are disjoint, as $\mathbb{P}[A \cap B] = 0$. But these events are not independent: $\mathbb{P}[B \mid A] = 0$, but $\mathbb{P}[B] = 1/6$.

Since disjoint events have $\mathbb{P}[A \cap B] = 0$, we can see that the only time when disjoint A and B are independent is when either $\mathbb{P}[A] = 0$ or $\mathbb{P}[B] = 0$.

2 Balls and Bins

Note 14

Suppose you throw b balls into n labeled bins one at a time.

(a) What is the probability that the first bin is empty?

(b) What is the probability that the first k bins are empty?

(c) Let A be the event that at least k bins are empty. Let m be the number of subsets of k bins out of the total n bins. If we assume A_i is the event that the ith subset of k bins is empty. Then we can write A as the union of A_i's:

$$A = \bigcup_{i=1}^{m} A_i.$$

Compute m in terms of n and k, and use the union bound to give an upper bound on the probability $\mathbb{P}[A]$.

(d) What is the probability that the second bin is empty given that the first one is empty?

(e) Are the events that "the first bin is empty" and "the first two bins are empty" independent?

(f) Are the events that "the first bin is empty" and "the second bin is empty" independent?

Solution: Since the balls are thrown one at a time, there is an ordering, and so we are sampling with replacement where order matters rather than where it doesn't (which would correspond to each configuration in the stars and bars setup being equally likely).

(a) Note that this is a uniform sample space, with outcomes representing all possible ways to throw each ball individually into the bins. Here, $|\Omega| = n^b$, as each of the b balls has n possible bins to fall into, and out of these possibilities, $(n-1)^b$ of them leave the first bin empty—each ball would then have $n-1$ possible bins to fall into. This gives us an overall probability $\left(\frac{n-1}{n}\right)^b$ that the first bin is empty.

Equivalently, we can note that each throw is independent of all of the other throws. Since the probability that ball i does not land in the first bin is $\frac{n-1}{n}$, the probability that all of the balls do not land in the first bin is $\left(\frac{n-1}{n}\right)^b$.

(b) Similar to the previous part, we have the same uniform sample space of size n^b. Now, there are a total of $(n-k)^b$ possible ways to throw the balls into bins such that the first k bins are empty—each ball has $n-k$ possible bins to fall into.

Alternatively, we can similarly make use of independence. Since the probability that ball i does not land in the first k bins is $\frac{n-k}{n}$, the probability that all of the balls do not land in the first k bins is $\left(\frac{n-k}{n}\right)^b$.

(c) We use the union bound. Then

$$\mathbb{P}[A] = \mathbb{P}\left[\bigcup_{i=1}^{m} A_i\right] \leq \sum_{i=1}^{m} \mathbb{P}[A_i].$$

We know the probability of the first k bins being empty from part (b), and this is true for any set of k bins, so

$$\mathbb{P}[A_i] = \left(\frac{n-k}{n}\right)^b.$$

Then,

$$\mathbb{P}[A] \leq m \cdot \left(\frac{n-k}{n}\right)^b = \binom{n}{k}\left(\frac{n-k}{n}\right)^b.$$

(d) Using Bayes' Rule:

$$\begin{aligned}
\mathbb{P}[\text{2nd bin empty} \mid \text{1st bin empty}] &= \frac{\mathbb{P}[\text{2nd bin empty} \cap \text{1st bin empty}]}{\mathbb{P}[\text{1st bin empty}]} \\
&= \frac{(n-2)^b/n^b}{(n-1)^b/n^b} \\
&= \left(\frac{n-2}{n-1}\right)^b
\end{aligned}$$

Alternate solution: We know bin 1 is empty, so each ball that we throw can land in one of the remaining $n-1$ bins. We want the probability that bin 2 is empty, which means that each ball cannot land in bin 2 either, leaving $n-2$ bins. Thus for each ball, the probability that bin 2 is empty given that bin 1 is empty is $\frac{n-2}{n-1}$. For b total balls, this probability is $\left(\frac{n-2}{n-1}\right)^b$.

(e) They are dependent. Knowing the latter means the former happens with probability 1.

(f) In part (c) we calculated the probability that the second bin is empty given that the first bin is empty: $\left(\frac{n-2}{n-1}\right)^b$. The probability that the second bin is empty (without any prior information) is $\left(\frac{n-1}{n}\right)^b$. Since these probabilities are not equal, the events are dependent.

3 Birthdays

Note 14 Suppose you record the birthdays of a large group of people, one at a time until you have found a match, i.e., a birthday that has already been recorded. (Assume there are 365 days in a year.)

(a) What is the probability that after the first 3 people's birthdays are recorded, no match has occurred (i.e. each person has a unique birthday)?

(b) What is the probability that the first 3 people all share the same birthday?

(c) What is the probability that it takes more than 20 people for a match to occur?

(d) What is the probability that it takes exactly 20 people for a match to occur?

(e) Suppose instead that you record the birthdays of a large group of people, one at a time, until you have found a person whose birthday matches your own birthday. What is the probability that it takes exactly 20 people for this to occur?

Solution:

(a) $\dfrac{364}{365} \cdot \dfrac{363}{365}$.

(b) $\left(\dfrac{1}{365}\right)^2$.

(c)

$$\begin{aligned}\mathbb{P}[\text{it takes more than 20 people}] &= \mathbb{P}[\text{20 people don't have the same birthday}] \\ &= \frac{365!/(365-20)!}{365^{20}} = \boxed{\frac{\mathbf{365!}}{\mathbf{345!365^{20}}}} \approx .589.\end{aligned}$$

Another explanation that does not use counting:

The first person can have any birthday. The second person must have a different birthday from the first person, which occurs with probability $364/365$. The third person must have a different birthday from the first two people, which occurs with probability $363/365$. Generalizing, the ith person must have a different birthday from the first $i-1$ people, which occurs with probability $(365-(i-1))/365$. Hence,

$$\begin{aligned}\mathbb{P}[\text{it takes more than 20 people}] &= \frac{365-19}{365} \times \frac{365-18}{365} \times \cdots \times \frac{363}{365} \times \frac{364}{365} \\ &\approx .589.\end{aligned}$$

(d) The probability that it takes exactly 20 people is the probability that the first 19 people have different birthdays **and** the 20th person shares a birthday with one of the first 19 people.

How many total ways can the birthdays be chosen for 20 people? 365^{20}.

How many ways can the birthdays be chosen so the first 19 have different birthdays and the 20^{th} person shares a birthday with the first 19? Well, the first person has 365 choices, the second has 364 choices left, and so on until the nineteenth person has $(365-19+1)=347$ choices left. Then, the 20^{th} person has 19 choices for his birthday. So in total, there are $365 \cdot 364 \cdots 348 \cdot 347 \cdot 19 = (365!/346!) \cdot 19$ ways of getting what we want. So

$$\mathbb{P}[\text{it takes exactly 20 people}] = \frac{365 \cdot 364 \cdots 348 \cdot 347 \cdot 19}{365^{20}} = \boxed{\mathbf{\frac{365! \cdot 19}{346! 365^{20}}}} \approx .032.$$

Another explanation that does not use counting:

As before, the ith person must have a different birthday from the first $i-1$ people, with probability $(365-(i-1))/365$, for $i=1,\ldots,19$. The 20th person must share a birthday with one of the first 19 people (who all have distinct birthdays), so the probabiltiy is $19/365$. Hence,

$$\begin{aligned}\mathbb{P}[\text{it takes exactly 20 people}] &= \frac{19}{365} \times \frac{365-18}{365} \times \cdots \times \frac{363}{365} \times \frac{364}{365} \\ &\approx .032.\end{aligned}$$

(e) The probability that it takes exactly 20 people is the probability that the first 19 people don't have your birthday and the 20th person has your birthday.

Similar to the last problem, there are 364 choices for the first person's birthday to be different than yours, 364 for the second person, and so on until the nineteenth person has 364 choices. Then, the 20^{th} person has exactly 1 choice to have your birthday. So the total number of ways to get what we want is $364^{19} \cdot 1$. There are 365^{20} possibilities total. So

$$\mathbb{P}[\text{it takes exactly 20 people}] = \boxed{\mathbf{\frac{364^{19}}{365^{20}}}} \approx .0026.$$

Another explanation that does not use counting:

Each of the 19 people who do not share your birthday do so with probability $364/365$, and the last person must share your birthday with probability $1/365$. Hence,

$$\begin{aligned}\mathbb{P}[\text{it takes exactly 20 people}] &= \frac{364}{365} \times \frac{364}{365} \times \cdots \times \frac{364}{365} \times \frac{1}{365} \\ &= \frac{364^{19} \times 1}{365^{20}} \\ &\approx 0.0026.\end{aligned}$$