Ma 3/103 KC Border
Introduction to Probability and Statistics Winter 2015

Chapter 4: Binomial Distribution; Bayes' Law

Relevant textbook passages:

Pitman [5]: Sections 1.4, 1.5, 2.1

Larsen–Marx [4]: Sections 2.4–2.7, 3.2

4.1 Bernoulli Trials

A **Bernoulli trial** is a random experiment with two possible outcomes, traditionally labeled "success" or 1 and "failure" or 0. The probability of success is traditionally denoted p.

Pitman [5]: p. 27

4.2 The Binomial Distribution

If there are n stochastically independent Bernoulli trials with the same probability p of success, the probability distribution of the number of successes is called the **Binomial distribution**. To get exactly k successes, there must be $n-k$ failures. There are

Pitman [5]: § 2.1

$$\binom{n}{k} = \frac{n!}{k!\,(n-k)!}$$

such outcomes (where by convention $0! = 1$), and by independence each has probability $p^k(1-p)^{n-k}$. Thus

$$P\,(k \text{ successes in } n \text{ independent Bernoulli trials}) = \binom{n}{k} p^k (1-p)^{n-k}.$$

Since $p + (1-p) = 1$ and $1^n = 1$, the Binomial Theorem assures us that the binomial distribution is a probability distribution.

4.2.1 Example (The probability of n heads in $2n$ coin flips) For a fair coin the probability of n heads in $2n$ coin flips is

$$\binom{2n}{n} \left(\frac{1}{2}\right)^{2n}.$$

We can see what happens to this for large n by using Stirling's approximation:

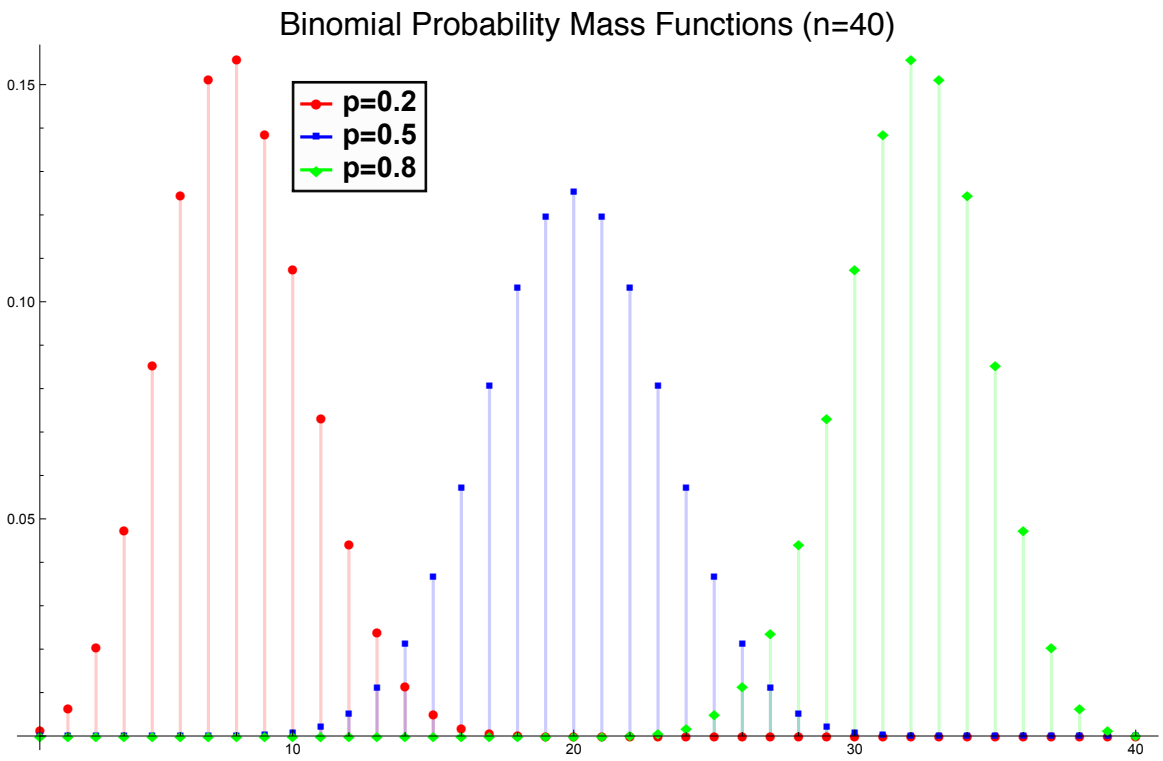


Figure 4.1. Binomial Probabilities

4.2.2 Proposition (Stirling's formula)

$$n! = e^{-n}n^n\sqrt{2\pi n}(1+\varepsilon_n)$$

where $\varepsilon_n \to 0$ *as* $n \to \infty$.

For a proof, see, e.g., Feller [2, p. 52] or Ash [1, pp.43–45], or the exercises in Pitman [5, p. 136].

Thus we may write

$$\frac{(2n)!}{n!\,n!} = \frac{e^{-2n}(2n)^{2n}\sqrt{4\pi n}}{e^{-n}e^{-n}n^nn^n\sqrt{2\pi n}\sqrt{2\pi n}}(1+\delta_n) = \frac{2^{2n}}{\sqrt{\pi n}}(1+\delta_n),$$

where $\delta_n \to 0$ as $n \to \infty$.

So the probability of n heads in $2n$ attempts is

$$\begin{aligned}&\frac{2^{2n}}{\sqrt{\pi n}}2^{-2n}(1+\delta_n)\\ &= \frac{1}{\sqrt{\pi n}}(1+\delta_n)\\ &\longrightarrow 0\end{aligned}$$

as $n \to \infty$.

What about the probability of between $n-k$ and $n+k$ heads in $2n$ tosses? Well the probability of getting j heads in $2n$ tosses is $\binom{2n}{j}(1/2)^{2n}$, and this is maximized at $j = n$ (See, e.g., Pitman [5, p. 86].) So we can use this as an upper bound. Thus for $k \geqslant 1$

$$\begin{aligned}P\left(\text{between } n-k \text{ and } n+k \text{ heads}\right) &< \frac{2k+1}{\sqrt{\pi n}}(1+\delta_n)\\ &\longrightarrow 0\end{aligned}$$

as $n \to \infty$.

So any reasonable "law of averages" will have to let k grow with n. We will come to this in a few more lectures.

□

4.3 The Multinomial Distribution

Larsen–Marx [4]: Section 10.2, pp. 494–499
Pitman [5]: p. 155

The **Multinomial distribution** generalizes the binomial distribution to random experiments with more than two outcomes. If there are m possible outcomes and the i^{th} outcome has probability p_i, then in n independent trials, if $k_1 + \cdots + k_m = n$,

$$P\left(k_i \text{ outcomes of type } i,\ i = 1, \ldots, m\right) = \frac{n!}{k_1! \cdot k_2! \cdot \cdots \cdot k_m} p_1^{k_1} \cdot p_2^{k_2} \cdot \cdots \cdot p_m^{k_m}.$$

4.3.1 Example Suppose you roll 9 dice. What is the probability of getting 3 aces (ones) and 6 boxcars (sixes)?

$$\frac{9!}{3!\,0!\,0!\,0!\,0!\,6!}\left(\frac{1}{6}\right)^9 = 94\frac{1}{10,077,696} \approx 0.0000083.$$

(Recall that $0! = 1$.) □

4.4 The Negative Binomial Distribution

Pitman [5]: p. 213
Larsen–Marx [4]: § 4.5

The **Negative Binomial Distribution** is the probability distribution of the number of independent trials need for a given number of heads. What is the probability that the r^{th} success occurs on trial t, for $t \geqslant r$?

For this to happen, there must be $t-r$ failures and $r-1$ successes in the first $t-1$ trials, with a success on trial t. By independence happens with the binomial probability for $r-1$ successes on $t-1$ trials times the probability p of success on trial t:

$$\binom{t-1}{r-1} p^r (1-p)^{t-r} \qquad (t \geqslant r).$$

Of course, the probability is 0 for $t < r$. The special case $r = 1$ (number of trials to the first success) is called the **Geometric Distribution**.

4.5 Independence and conditional probability

If A and B are independent, then

$$P(B \,\big|\, A) = \frac{P(BA)}{P(A)} = \frac{P(B) \cdot P(A)}{P(A)} = P(B)$$

or

$$P(AB) = P(A)P(B).$$

4.5.1 Product sample spaces

Consider two random experiments $(S_1, \mathcal{E}_1, P_1)$ and $(S_2, \mathcal{E}_2, P_2)$. Intuitively the outcome of each experiment is independent (in the nontechnical sense) of the other, then the "joint experiment" has sample space $S_1 \times S_2$, the Cartesian product of the two sample spaces. The set of events will certainly include sets of the form $E_1 \times E_2$ where $E_i \in \mathcal{E}_i$, but will also include just enough other sets to make a σ-algebra of events. If the outcomes are independent, we expect that the probability

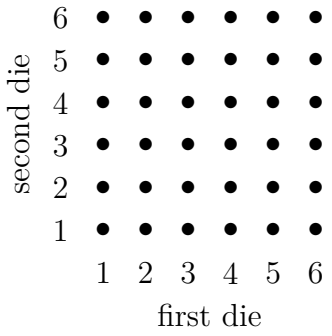
$$P(E_1 \times E_2) = P\left((E_1 \times S_2) \bigcap (S_1 \times E_2)\right) = P_1(E_1) \times P_2(E_2).$$

In this case, the events $(E_1 \times S_2)$ and $(S_1 \times E_2)$ in the joint experiment (which are essentially E_1 in experiment 1 and E_2 in experiment 2) are technically **stochastically independent**. This "product structure" is typical (but not definitional) of stochastic independence.

We usually think that successive coin tosses, rolls of dice, etc., are independent. As an example of experiments that are not independent, consider testing potentially fatal drugs on lab rats, with the same set of rats. If a rat dies in the first experiment, it diminishes the probability he survives in the second.

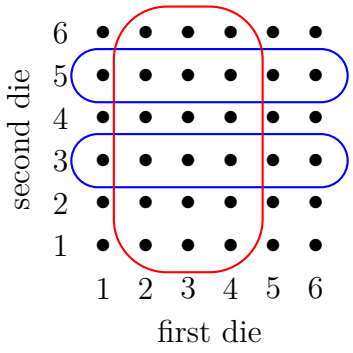
4.5.2 An example

Consider the random experiment of independently rolling two dice. There are 36 equally likely outcomes and the sample space S can be represented by the following rectangular array.



The assumption that each outcome is equally likely amounts to assuming that the probability of the product of an event in terms of the first die and one in terms of the second die is the product of the probabilities of each event.

Consider the event E that the second die is 3 or 5, which contains 12 points; and the event F that the first is 2, 3, or 4, which contains 18 points. Thus $P(E) = 12/36 = 1/3$, and $P(F) = 18/36 = 1/2$.

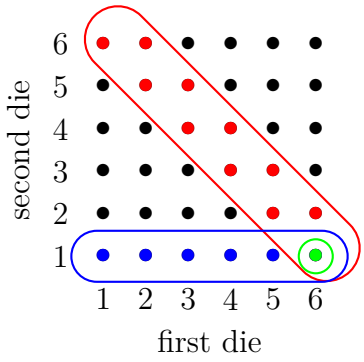


The intersection has 6 points, so $P(EF) = 6/36 = 1/6$. Observe that

$$P(EF) = \frac{1}{6} = \frac{1}{3}\frac{1}{2} = P(E)P(F).$$

4.5.3 Another example

The red oval represent the event A that the sum of the two dice is 7 or 8. It contains 11 points, so it has probability 11/36. Let B be the event that the second die is 1. It is outlined in blue, and has 6 points, and so has probability $6/36 = 1/6$. The event BA is circled in green, and consists of the single point $(6, 1)$.



If we **"condition" on A**, we can ask, what is the probability of the event B = (the second die is 1) **given** that we know that A has occurred, denoted $B \,\big|\, A$. Thus

$$P(B \,\big|\, A) = \frac{P(BA)}{P(A)} = \frac{1/36}{11/36} = \frac{1}{11}.$$

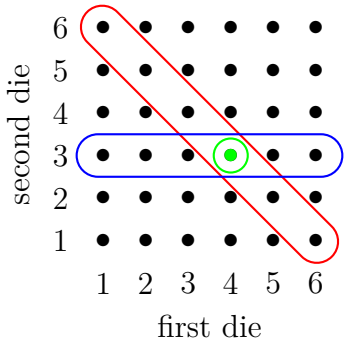
That is, we count the number of points in BA and divide by the number of points in A.

Similarly

$$P(A \,\big|\, B) = \frac{P(AB)}{P(B)} = \frac{1/36}{6/36} = \frac{1}{6}.$$

4.5.4 Yet another example

To see that not every example of independence involves "orthogonal" product sets, consider this example involving rolling two dice independently. Event A is the event that the sum is 7, and event B is the event that the second die is 3. These events are not of the form $E_i \times S_j$, yet they are stochastically independent.



4.5.5 Conditional probability, continued

We can think of $P(\cdot \,\big|\, A)$ as a "renormalized" probability, with A as the new sample space. That is,

1. $P(A \,\big|\, A) = 1$.

2. If $BC = \varnothing$, then
$$P(B \cup C \,\big|\, A) = P(B \,\big|\, A) + P(C \,\big|\, A)$$

3. $P(\varnothing \,\big|\, A) = 0$.

Proof of (2):
$$\begin{aligned} P(B \cup C \,\big|\, A) &= \frac{P\left((B \cup C)A\right)}{P(A)} \\ &= \frac{P\left((BA) \cup (CA)\right)}{P(A)} \\ &= \frac{P(BA) + P(CA)}{P(A)} \\ &= \frac{P(BA)}{P(A)} + \frac{P(CA)}{P(A)} \\ &= P(B \,\big|\, A) + P(C \,\big|\, A). \end{aligned}$$

4.6 Bayes' Rule

Now
$$P(B \,\big|\, A) = \frac{P(BA)}{P(A)}, \qquad P(A \,\big|\, B) = \frac{P(AB)}{P(B)},$$
so we have the **Multiplication Rule**
$$P(AB) = P(B \,\big|\, A) \cdot P(A) = P(A \,\big|\, B) \cdot P(B).$$
and

4.6.1 Bayes' Rule
$$P(B \,\big|\, A) = P(A \,\big|\, B) \frac{P(B)}{P(A)}.$$

We can also discuss odds using Bayes' Law. Recall that the odds against B are $P(B^c)/P(B)$. Now suppose we know that event A has occurred. The **posterior odds** against B are now
$$\frac{P(B^c \,\big|\, A)}{P(B \,\big|\, A)} = \frac{P(A \,\big|\, B^c) \frac{P(B^c)}{P(A)}}{P(A \,\big|\, B) \frac{P(B)}{P(A)}}. = \frac{P(A \,\big|\, B^c)}{P(A \,\big|\, B)} \frac{P(B^c)}{P(B)}.$$

The term $P(B^c)/P(B)$ is the **prior odds** ratio. Now let's compare the posterior and prior odds:

$$\frac{P(B^c \,\big|\, A)/P(B \,\big|\, A)}{P(B^c)/P(B)} = \frac{P(A \,\big|\, B^c)}{P(A \,\big|\, B)}.$$

The right-hand side term $\dfrac{P(A \,\big|\, B^c)}{P(A \,\big|\, B)}$ is called the **likelihood ratio** or the **Bayes factor**.

Aside: According to my favorite authority on such matters, the 13th edition of the *Chicago Manual of Styles* [6], we should write Bayes's Rule, but nobody does.

4.6.2 Proposition (Law of Average Conditional Probability) *[Cf. Pitman [5], § 1.4, p. 41.] Let $B_1, \ldots, B_n$ be a **partition** of S. Then for any $A \in \mathcal{E}$,*

$$P(A) = P(A \,\big|\, B_1)P(B_1) + \cdots + P(A \,\big|\, B_n)P(B_n).$$

Proof: This follows from the fact than for each i, $P(A \,\big|\, B_i)P(B_i) = P(AB_i)$ and the fact that since the B_is partition S, we have

$$A = \bigcup_{i=1}^{n} (AB_i),$$

and for $i \neq j$, $(AB_i)(AB_j) = \varnothing$. Now just use additivity of P. ∎

We can use this to rephrase Bayes' Rule as follows.

Pitman [5]: p. 49

4.6.3 Theorem (Bayes' Rule) *Let the events $B_1, \ldots, B_n$ be a partition of S. The for any event A and any i*

$$P(B_i \,\big|\, A) = \frac{P(A \,\big|\, B_i)P(B_i)}{P(A \,\big|\, B_1)P(B_1) + \cdots + P(A \,\big|\, B_n)P(B_n)}$$

4.6.4 Example (Card placement) Consider a well-shuffled standard deck of 52 cards. By well-shuffled I mean that every ordering of the cards is equally likely. (Note that there are $52! \approx 8 \times 10^{67}$ orderings, so it is not possible to list them all.) In that case, the probability that the top card is an ace is equal to the probability that the second card is an ace and both are equal to $4/52 = 1/13$. (We'll come back to this later on.)

I was asked how this can be so, since if the first card is an ace, the probability that the second card is an ace is only 3/51. This is indeed true, but so what? Let

A denote the event the top card is an ace,

and let

B denote the event that the second card is an ace.

Then I claim that $P(A) = P(B) = 4/52 = 1/13$, and also $P(B \mid A) = 3/51$. By the above proposition,

$$\begin{aligned} P(B) &= \underset{\frac{3}{51}}{P(B \mid A)}\,\underset{\frac{1}{13}}{P(A)} + \underset{\frac{4}{51}}{(B \mid A^{\mathrm{c}})}\,\underset{\frac{12}{13}}{P(A^{\mathrm{c}})} \\ &= \frac{3}{51} \cdot \frac{1}{13} + \frac{4}{51} \cdot \frac{12}{13} \\ &= \frac{3 + 48}{13 \times 51} \\ &= \frac{1}{13} \end{aligned}$$

□

4.7 Bayes' Law and False Positives

Recall Bayes' Rule:

$$P(B \mid A) = \frac{P(A \mid B)P(B)}{P(A \mid B)P(B) + P(A \mid B^{\mathrm{c}})P(B^{\mathrm{c}})}$$

"When you hear hoofbeats, think horses, not zebras," is advice commonly given to North American medical students. It means that when you are presented with an array of symptoms, you should think of the most likely, not the most exotic explanation.

Pitman [5]: p. 52

4.7.1 Example (Hoofbeats) Suppose there is a diagnostic test, e.g., CEA (carcinoembryonic antigen) levels, for a particular disease (e.g., colon cancer or rheumatoid arthritis). It is in the nature of human physiology and medical tests that they are imperfect. That is, you may have the disease and the test may not catch it, or you may not have the disease, but the test will suggest that you do. (These mistakes are usually labeled, unimaginatively and opaquely, as Type 1 and Type 2 errors.)

Suppose further that in fact, one in a hundred people suffer from disease D. Suppose that the test is accurate in the sense that if you have the disease, it catches it (tests positive) 99% of the time. But suppose also that in one case in a hundred, it reports falsely that a someone has the disease when in fact they do not.

What is the probability that a randomly selected individual who is tested and has a positive test result actually has the disease? What is the probability that someone who tests negative for the disease actually is disease free?

Let D denote the event that a randomly selected person has the disease, and $\neg D$ be the event the person does not have the disease. Let $+$ denote the event

that the test is positive, and $-$ be the event the test is negative. We are told

$$P(D) = 0.01, \quad P(\neg D) = 0.99,$$

$$\text{Prob}\left(+ \,\middle|\, D\right) = 0.99 \text{ and Prob}\left(+ \,\middle|\, \neg D\right) = 0.01,$$

so

$$\text{Prob}\left(- \,\middle|\, D\right) = 0.01 \text{ and Prob}\left(- \,\middle|\, \neg D\right) = 0.99,$$

For the first question we want to know $P(D \,|\, +)$. By Bayes' Rule,

$$\begin{aligned} P(D \,|\, +) &= \frac{P(+ \,|\, D)P(D)}{P(+ \,|\, D)P(D) + P(+ \,|\, \neg D)P(\neg D)} \\ &= \frac{0.99 \times 0.01}{(0.99 \times 0.01) + (0.01 \times 0.99)} \\ &= 0.5. \end{aligned}$$

In other words, if the test reports that you have the disease there is only a fifty-fifty chance that you do have the disease.

For the second question, we want to know $P(\neg D \,|\, -)$, which is

$$\begin{aligned} P(\neg D \,|\, -) &= \frac{P(- \,|\, \neg D)P(\neg D)}{P(- \,|\, \neg D)P(\neg D) + P(- \,|\, D)P(D)} \\ &= \frac{0.99 \times 0.99}{(0.99 \times 0.99) + (0.01 \times 0.01)} \\ &= 0.999898. \end{aligned}$$

That is, a negative result means it is very unlikely you do have the disease. □

4.7.1 Hoofbeats revisited

There is something wrong with the hoofbeat example. The conclusions would be correct if the person being tested were randomly selected from the population at large. But this is rarely the case. Usually someone is tested only if there are symptoms or other reasons to believe the person may have the disease or have been exposed to it.

So let's modify this example by introducing a symptom S. Suppose that if you have the disease there is a 80% chance you have the symptom, but if you do not have the disease, there is only a 30% chance you have the symptom. Suppose further that only those exhibiting the symptom are tested. Now the probability of having the disease given a positive test is really the probability of having the disease given the symptom *and* a positive test result.

$$P\left(D \,\middle|\, +, S\right) = \frac{P\left(D, +, S\right)}{P\left(+, S\right)} = \frac{P\left(D \,\middle|\, +, S\right) P\left(+ \,\middle|\, S\right) P\left(S\right)}{P\left(+ \,\middle|\, S\right) P\left(S\right)}$$

4.8 A family example

Assume that the probability that a child is male or female is 1/2, and that the sex of children in a family are independent trials. So for a family with two children the sample space is

$$S = \{(F,F),(F,M),(M,F),(M,M)\},$$

and each outcome has probability 1/4. Now suppose you are informed that the family has at least one girl. What is the probability that the other child is a boy? Let

$$G = \{(F,F),(F,M),(M,F)\}$$

be the event that there is at least one girl. The event that "the other child is a boy" corresponds to the event

$$B = \{(F,M),(M,F)\}.$$

The probability $P(B \,\big|\, G)$ is thus 2/3.

One year a student asked "How does knowing a family has a girl make it more likely to have a boy?" It doesn't. The probability that the family has a boy is not 1/2. It's actually 3/4. So learning that one child is a girl reduces the probability of at least one boy from 3/4 to 2/3.

Now suppose you are told that the eldest child is a girl. This is the event

$$E = \{(F,F),(F,M)\}.$$

Now the probability that the other child is a boy is 1/2.

This means that the information that "there is at least one girl" and the information that "the eldest is a girl" are really different pieces of information. While it might seem that birth order is irrelevant, a careful examination of the outcome space shows that it is not.

Another variant is this. Suppose you meet girl X, who announces "I have one sibling." What is the probability that it is a boy?

The outcome space here is not obvious. I argue that it is:

$$\{(X,F),(X,M),(F,X),(M,X)\}.$$

We don't know the probabilities of the individual outcomes, but (X,F) and (X,M) are equally likely, and (F,X) and (M,X) are equally likely. Let

$$P\{(X,F)\} = P\{(X,M)\} = a \quad \text{and} \quad P\{(F,X)\} = P\{(M,X)\} = b.$$

Thus $2a + 2b = 1$, so $a + b = 1/2$. The probability of X's sibling being a boy is

$$P\{(X,M),(M,X)\} = P\{(X,M)\} + P\{(M,X)\} = a + b = 1/2.$$

4.9 Conditioning and intersections

We already know that $P(AB) = P(A \,\big|\, B)P(B)$. This extends by a simple induction argument to the following (Pitman [5, p. 56]):

$$P(A_1A_2\cdots A_n) = P(A_n \,\big|\, A_{n-1}\cdots A_1)P(A_{n-1}\cdots A_1)$$

but

$$P(A_{n-1}\cdots A_1) = P(A_{n-1} \,\big|\, A_{n-2}\cdots A_1)P(A_{n-2}\cdots A_1),$$

so continuing in this fashion and using the more compact notation for intersection, we obtain

$$\begin{aligned}&P(A_1A_2\cdots A_n) =\\ &\quad P(A_n \,\big|\, A_{n-1}\cdots A_1)P(A_{n-1} \,\big|\, A_{n-2}\cdots A_1)\cdots P(A_3 \,\big|\, A_2A_1)P(A_2 \,\big|\, A_1)P(A_1)\end{aligned}$$

4.10 The famous birthday problem

Pitman [5]: § pp. 62–63

Assume that there are only 365 possible birthdays, all equally likely,[1] and assume that in a typical group they are stochastically independent.[2] In a group of size $n \leqslant 365$, what is the probability that at least two people share a birthday? The sample space for this experiment is

$$S = \{1, \ldots, 365\}^n,$$

which gets big fast. ($|S|$ is about 1.7×10^{51} when $n = 20$.)

This is a problem where it is easier to compute the complementary probability, that is, the probability that all the birthdays are distinct. Number the people from 1 to n. Let A_k be the event that the birthdays of persons 1 through k are distinct. (Note that $A_1 = S$.)

Observe that

$$A_{k+1} \subset A_k$$

for every k, which means $A_k = A_kA_{k-1}\cdots A_1$ for every k. Thus

$$P(A_{k+1} \,\big|\, A_kA_{k-1}\cdots A_1) = P(A_{k+1} \,\big|\, A_k).$$

[1] It is unlikely that all birthdays are equally likely. For instance, it was reported that many mothers scheduled C-sections so their children would be born on 12/12/2012. There are also special dates, such as New Year's Eve, on which children are more likely to be conceived. It also matters how the group is selected. Malcolm Gladwell [3, pp. 22–23] reports that a Canadian psychologist named Roger Barnsley discovered the "iron law of Canadian hockey: in *any* elite group of hockey players—the very best of the best—40% of the players will be born between January and March, 30% between April and June, 20% between July and September, and 10% between October and December." Can you think of an explanation for this?

[2] If this were a gathering of identical twins, the stochastic independence assumption would have to be jettisoned.

The formula for the probability of an intersection in terms of conditional probabilities implies

$$\begin{aligned} P(A_n) &= P(A_1 \cdots A_n) \\ &= P(A_n \,\big|\, A_{n-1} \cdots A_1) P(A_{n-1} \,\big|\, A_{n-2} \cdots A_1) \cdots P(A_2 \,\big|\, A_1) P(A_1) \\ &= P(A_n \,\big|\, A_{n-1}) P(A_{n-1} \,\big|\, A_{n-2}) \cdots P(A_2 \,\big|\, A_1) P(A_1) \end{aligned}$$

4.10.1 Claim *For $k < 365$,*

$$P(A_{k+1} \,\big|\, A_k) = \frac{365 - k}{365}.$$

While in many ways this claim is obvious, let's plug and chug.

Proof: By definition,

$$P(A_{k+1} \,\big|\, A_k) = \frac{P(A_{k+1} A_k)}{P(A_k)} = \frac{P(A_{k+1})}{P(A_k)}.$$

Now there are 365^k equally likely possible lists of birthdays for the k people, since it is quite possible to repeat a birthday. How many give distinct birthdays? There are 365 possibilities for the first person, but after that only 364 choices remain for the second, etc. Thus there are $365!/(365 - k)!$ lists of distinct birthdays for k people. So for each $k \leqslant 365$,

$$P(A_k) = \frac{365!}{(365 - k)!\, 365^k},$$

which in turn implies

$$P(A_{k+1} \,\big|\, A_k) = \frac{P(A_{k+1})}{P(A_k)} = \frac{\frac{365!}{(365-k-1)!\, 365^{k+1}}}{\frac{365!}{(365-k)!\, 365^k}} = \frac{365 - k}{365},$$

as claimed. ∎

Thus

$$P(A_n) = \prod_{k=1}^{n-1} \frac{365 - k}{365}.$$

The probability that at least two share a birthday is 1 minus this product. Here is some Mathematica code to make a table

```
TableForm[
 Table[{n, N[1 - Product[(365 - k)/365, {k, 1, n - 1}]]}, {n, 20, 30}]
 ]  // TeXForm
```

to produce this table:[3]

n	Prob. of sharing
20	0.411438
21	0.443688
22	0.475695
23	0.507297
24	0.538344
25	0.5687
26	0.598241
27	0.626859
28	0.654461
29	0.680969
30	0.706316

Pitman [5, p. 63] gets 0.506 for $n = 23$, but Mathematica gets 0.507. Hmmm.

Bibliography

[1] R. B. Ash. 2008. *Basic probability theory*. Mineola, New York: Dover. Reprint of the 1970 edition published by John Wiley and Sons.

[2] W. Feller. 1968. *An introduction to probability theory and its applications*, 3d. ed., volume 1. New York: Wiley.

[3] M. Gladwell. 2008. *Outliers: The story of success*. New York, Boston, London: Little, Brown.

[4] R. J. Larsen and M. L. Marx. 2012. *An introduction to mathematical statistics and its applications*, fifth ed. Boston: Prentice Hall.

[5] J. Pitman. 1993. *Probability*. Springer Texts in Statistics. New York, Berlin, and Heidelberg: Springer.

[6] University of Chicago Press Editorial Staff, ed. 1982. *The Chicago manual of style*, 13th ed. Chicago: University of Chicago Press.

[3] I did edit the TeX code to make the table look better.