Computing the Probability of an Event

Some probabilities can be computed directly. For example, if you're dealing with a collection of equiprobable possibilities, mere counting might give you an answer.

Example: What is the probability that a randomly dealt five-card poker hand (i.e., all possible hands are equally likely) is a "full house," i.e., contains three cards of one rank and two cards of another?

There are 13 ways to choose the rank of the three identical cards, and 4 ways to choose which three of the four cards of that rank will be held. Then there remain 12 ways to choose the rank of the pair, and 6 ways to choose some specific two cards of that rank. Therefore, there are $13 \cdot 4 \cdot 12 \cdot 6 = 3{,}744$ different full houses, out of the 2,598,560 possible hands: The probability is $3{,}744 / 2{,}598{,}560 = 0.144\%$.

Quite often, however, a direct way to compute a desired probability isn't immediately apparent. In such cases, it may be productive to either recast or decompose the original problem. One simple recasting is to seek the probability of the complementary event. Frequently, "at least" problems can be solved this way, by flipping them into "at most" problems.

Example: What's the probability of getting at least one head in the course of 5 coin tosses?

One could work out all the separate cases – one head in five tosses, two heads, or three or four or five. But it's much easier to find the probability of no heads, and subtract that from one:

$$\text{Pr(at least one head in 5 tosses)}$$
$$= 1 - \text{Pr(no heads in 5 tosses)} = 1 - (0.5)^5 = 96.9\%.$$

Example: What's the probability that at least two people in a group of 5 were born in the same month?

If we try to work this out directly, we need to consider many different possibilities: Two born in the same month and the others in different months, or three born in the same month, or a pair born during one month and another pair during another month. But it's easier to find the probability that all five were born in different months, and subtract that from one:

$$1 - (12/12) \cdot (11/12) \cdot (10/12) \cdot (9/12) \cdot (8/12) = 61.8\%.$$

If you're dealing with the conjunction of several events (an *"and"* problem) and the events are independent, just multiply their probabilities together. More generally, if there's some way to consider the events as occurring in sequence, try using this:

$$\text{Pr}(\, A_1 \text{ and } A_2 \text{ and } \dots \text{ } A_k \,) = \text{Pr}(\, A_1 \,) \cdot \text{Pr}(\, A_2 \mid A_1 \,) \cdot \dots \cdot \text{Pr}(\, A_k \mid A_1 \text{ and } A_2 \text{ and } \dots \text{ } A_{k-1} \,)\,.$$

Example: A purchasing agent is evaluating potential suppliers of an important electrical component for one of her company's products. There are eight firms that make components possibly fitting her company's needs. Her plan is to visit the firms in sequence, together with a company engineer, to test the components. When she finds a supplier with a satisfactory component, she will place an order. Actually, only three of the eight firms make satisfactory components. How likely is it that she will have to visit precisely four firms (i.e., that the first three visited will offer unsatisfactory components, and the fourth a satisfactory one)?

In order for her to need four visits, the outcomes of the four visits must be bad-bad-bad-good, in order. The sought-for probability is then

Pr(1st visit bad *and* 2nd visit bad *and* 3rd visit bad *and* 4th visit good)

$= $ Pr(1st visit bad) $\cdot$ Pr(2nd visit bad | 1st visit bad)
$\cdot$ Pr(3rd visit bad | 1st *and* 2nd visits bad)
$\cdot$ Pr(4th visit good | 1st *and* 2nd *and* 3rd visits bad)

$= 5/8 \cdot 4/7 \cdot 3/6 \cdot 3/5 = 10.7\%$.

If you can't break the problem down completely, a single step of this type might at least recast the problem.

Example: Assume the purchasing agent wishes to dual-source her order. What is the probability that it takes her four visits to find two good suppliers?

Pr(second good supplier found on 4th visit)

$= $ Pr(1 good supplier found in first 3 visits *and* another good supplier found on 4th visit)

$= $ Pr(exactly 1 good supplier found in first 3 visits)
 $\cdot$ Pr(a good supplier found on 4th visit | exactly 1 good supplier found in first 3 visits)

$$= \frac{\binom{3}{1} \cdot \binom{5}{2}}{\binom{8}{3}} \cdot \frac{2}{5} = 21.4\% \ .$$

Perhaps the event under investigation can be written as the disjunction of several events (an *"or"* problem). If the events are disjoint, find the probability of each separately and add the results.

Example: The Chicago Cubs have three games left against the Houston Astros, and two against the St. Louis Cardinals. They need to win at least four of these games to clinch a playoff berth. Given how they've fared against these opponents earlier in the season, you feel there's a 50% chance of them winning any given game against the Astros, and a 55% chance of them winning any given game against the Cardinals. How likely is it that the Cubs will make the playoffs?

Pr(Cubs make playoffs)

= Pr(Cubs win two against the Astros and two against the Cardinals
　　or Cubs win three against the Astros and one against the Cardinals
　　or Cubs win all five games)

= Pr(Cubs win two against the Astros and two against the Cardinals)
+ Pr(Cubs win three against the Astros and one against the Cardinals)
+ Pr(Cubs win all five games)

$= [3 \cdot (0.5)^2 \cdot (0.5)] \cdot (0.55)^2 + (0.5)^3 \cdot [2 \cdot (0.55) \cdot (0.45)] + (0.5)^3 \cdot (0.55)^2 = 21.3\%$.

However, if the event is the disjunction of events that might sometimes occur together, consider recasting it as an *"and"* problem by taking one of the following two approaches:

$$\textbf{Pr(A } \textit{or} \textbf{ B) = Pr(A) + Pr(B) } - \textbf{Pr(A } \textit{and} \textbf{ B) , or}$$

$$\textbf{Pr(A } \textit{or} \textbf{ B) = 1 } - \textbf{Pr(not-A } \textit{and} \textbf{ not-B) .}$$

Example: The Cubs will back into a post-season playoff berth if the Astros lose the final game of the season to the Reds (a 40% chance) or if the Cardinals lose their final game against the Pirates (a 45% chance). What is the probability that the Cubs will reach the playoffs?

Pr(Cubs reach the playoffs)

= Pr(Astros lose *or* Cardinals lose)

= Pr(Astros lose) + Pr(Cardinals lose) − Pr(Astros lose *and* Cardinals lose)

$= 0.40 + 0.45 - 0.40 \cdot 0.45 = 67\%$

= 1 − Pr(Astros win *and* Cardinals win)

$= 1 - 0.60 \cdot 0.55 = 67\%$.

Example: McDonald's prints 10,000 of each of 12 different types of stamps (each marked with a Zodiac sign). A particular pair of stamps (Pisces and Aquarius) can be turned in for a free Coca-Cola. If you collect a total of 8 stamps, what is the probability that you will be able to claim a free Coke?

This sounds like an "at least" problem, since we want the probability of winning at least one Coke. Let's try the complementary problem: What's the probability of not winning a Coke?

Pr(not winning a Coke) = Pr(no Pisces stamps *or* no Aquarius stamps)

= Pr(no Pisces stamps) + Pr(no Aquarius stamps)

– Pr(no Pisces stamps *and* no Aquarius stamps)

$$= \frac{\binom{110{,}000}{8}}{\binom{120{,}000}{8}} + \frac{\binom{110{,}000}{8}}{\binom{120{,}000}{8}} - \frac{\binom{100{,}000}{8}}{\binom{120{,}000}{8}} = 76.45\% \ .$$

So, the chance of winning a Coke is 23.55% .

If you seek a conditional probability, you can use the direct definition:

$$\mathbf{Pr(\ B \mid A\ ) = Pr(\ B \ \textit{and}\ A\ )\ /\ Pr(\ A\ )\ .}$$

Example: Three cards are placed into a bag. One card is red on both sides, another is blue on both sides, and the third is red on one side and blue on the other. You slide a card, selected at random, from the bag onto a table, and see that the side showing is red. What is the probability that the face-down side is also red?

Pr(face-down side is red | top side is red)

= Pr(a random card has both sides red) / Pr(a random card- face is red)

= (1/3) / (1/2) = 2/3 .

Alternatively, if Pr(A | B) would be easier to determine (which is frequently the case if the event B occurs earlier in time than does the event A), Bayes' Rule (or equivalently, a probability tree) might be useful:

$$\mathbf{Pr(\ B \mid A\ ) = Pr(\ A \mid B\ ) \cdot Pr(\ B\ )\ /\ Pr(\ A\ )\ .}$$

If your first thought upon encountering a problem is, "Well, it depends ... ," then a single event might be advantageously broken into disjoint events along the "it depends" dimension. If D₁, …, Dₖ are disjoint events which together cover all possibilities, then

$$\text{Pr}(\ C\ ) = \text{Pr}(\ C\ and\ D_1\ ) + … + \text{Pr}(\ C\ and\ D_k\ )$$
$$= \text{Pr}(\ C\ |\ D_1\ ) \cdot \text{Pr}(\ D_1\ ) + … + \text{Pr}(\ C\ |\ D_k\ ) \cdot \text{Pr}(\ D_k\ ) .$$

This "decomposition" method is often used to determine the denominator in calculations based on Bayes' Rule.

Example: A test well drilled on a tract of oil-bearing land will strike oil about 25% of the time. (Of course, if the well is drilled on non-bearing land, it will surely come up dry.) A preliminary seismic survey has indicated a 40% chance of oil being present on a particular tract. A test well is drilled, and comes up dry. What now is the chance that oil is indeed present?

Pr(test well comes up dry)

= Pr(test well would come up dry | oil is present) · Pr(oil is present)
  + Pr(test well would come up dry | oil is not present) · Pr(oil is not present)

= 0.75 · 0.4 + 1 · 0.6 = 90% .

Nature makes the "is there oil?" decision before we drill the test well: Try Bayes' Rule.

Pr(oil is present | test well is dry)

= Pr(test well is dry | oil is present) · Pr(oil is present) / Pr(test well is dry)

= 0.75 · 0.4 / 0.9 = 33.3% .

Finally, some problems can be systematically "stepped down" a notch. This might yield a recursive solution to the original problem.

Example: For another promotion, McDonald's cuts a deal with Michael Jordan. They will print a large (and equal) number of each of twelve "action photo" trading cards showing highlights of Jordan's basketball career. If you get a single randomly-selected card each time you eat at McDonald's, how many meals would you need to buy to have at least a 50% chance of receiving all 12 cards?

There are only two ways to end up with k distinct cards after n have been collected: Have k–1 distinct cards amongst the first n–1 and then get a new type, or have k distinct cards amongst the first n–1 and then get a duplicate.

Pr(k distinct cards amongst n)

= Pr(k–1 distinct cards amongst first n–1)
        · Pr(obtaining new type on n^{th} try | already hold k–1 distinct cards)
  + Pr(k distinct cards amongst first n–1)
        · Pr(obtaining a duplicate type on n^{th} try | already hold k distinct cards)

= Pr(k–1 distinct cards amongst first n–1) · (13–k) / 12
  + Pr(k distinct cards amongst first n–1) · k / 12 .

Since Pr(only a single type of card is held after n are collected) = $12 \cdot (1/12)^n = 1 / 12^{n-1}$, we know the values which go into the first column of the table below. The values that go into the first row are obvious. From these, the relationship above can be used to fill in the rest of the table. (Each table entry depends on the entry above it, and the entry above and to the left of it.)

		distinct cards collected											
		1	2	3	4	5	6	7	8	9	10	11	12
	1	100.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
	2	8.33%	91.67%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
	3	0.69%	22.92%	76.39%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
number	4	0.06%	4.46%	38.19%	57.29%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
of	5	0.00%	0.80%	13.26%	47.74%	38.19%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
cards	6	0.00%	0.14%	3.98%	25.86%	47.74%	22.28%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
collected	7	0.00%	0.02%	1.11%	11.60%	37.13%	38.99%	11.14%	0.00%	0.00%	0.00%	0.00%	0.00%
(i.e., meals eaten)	8	0.00%	0.00%	0.30%	4.70%	23.21%	41.16%	25.99%	4.64%	0.00%	0.00%	0.00%	0.00%
	9	0.00%	0.00%	0.08%	1.79%	12.80%	34.12%	35.74%	13.93%	1.55%	0.00%	0.00%	0.00%
	10	0.00%	0.00%	0.02%	0.65%	6.53%	24.53%	37.91%	24.18%	5.80%	0.39%	0.00%	0.00%
	11	0.00%	0.00%	0.01%	0.23%	3.16%	16.07%	34.38%	31.91%	12.41%	1.77%	0.06%	0.00%
	12	0.00%	0.00%	0.00%	0.08%	1.47%	9.88%	28.09%	35.60%	19.94%	4.58%	0.35%	0.01%
	13	0.00%	0.00%	0.00%	0.03%	0.67%	5.80%	21.32%	35.44%	26.82%	8.80%	1.09%	0.03%
	14	0.00%	0.00%	0.00%	0.01%	0.30%	3.29%	15.34%	32.51%	31.93%	14.04%	2.46%	0.13%
	15	0.00%	0.00%	0.00%	0.00%	0.13%	1.82%	10.59%	28.06%	34.78%	19.68%	4.60%	0.33%
	16	0.00%	0.00%	0.00%	0.00%	0.06%	0.98%	7.09%	23.12%	35.44%	25.10%	7.50%	0.71%
	17	0.00%	0.00%	0.00%	0.00%	0.02%	0.52%	4.63%	18.37%	34.29%	29.78%	11.06%	1.34%
	18	0.00%	0.00%	0.00%	0.00%	0.01%	0.28%	2.96%	14.17%	31.84%	33.39%	15.10%	2.26%
	19	0.00%	0.00%	0.00%	0.00%	0.00%	0.14%	1.87%	10.68%	28.60%	35.78%	19.40%	3.52%
	20	0.00%	0.00%	0.00%	0.00%	0.00%	0.07%	1.16%	7.90%	25.01%	36.97%	23.75%	5.14%
	21	0.00%	0.00%	0.00%	0.00%	0.00%	0.04%	0.71%	5.75%	21.39%	37.06%	27.93%	7.11%
	22	0.00%	0.00%	0.00%	0.00%	0.00%	0.02%	0.44%	4.13%	17.96%	36.23%	31.78%	9.44%
	23	0.00%	0.00%	0.00%	0.00%	0.00%	0.01%	0.26%	2.93%	14.85%	34.68%	35.17%	12.09%
	24	0.00%	0.00%	0.00%	0.00%	0.00%	0.01%	0.16%	2.07%	12.11%	32.61%	38.02%	15.02%
	25	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.10%	1.44%	9.77%	30.21%	40.29%	18.19%
	26	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.06%	1.00%	7.81%	27.62%	41.96%	21.55%
	27	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.03%	0.69%	6.19%	24.97%	43.07%	25.04%
	28	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.02%	0.48%	4.88%	22.35%	43.64%	28.63%
	29	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.01%	0.33%	3.82%	19.85%	43.73%	32.27%
	30	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.01%	0.22%	2.97%	17.49%	43.39%	35.91%
	31	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.15%	2.30%	15.32%	42.69%	39.53%
	32	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.10%	1.78%	13.34%	41.69%	43.09%
	33	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.07%	1.37%	11.56%	40.44%	46.56%
	34	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.05%	1.05%	9.98%	39.00%	49.93%
	35	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.03%	0.80%	8.58%	37.41%	53.18%
	36	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.02%	0.61%	7.35%	35.72%	56.30%

[*Challenge problem:* Eight cars pull out of a restaurant parking lot at roughly the same time. Due to on-going roadwork, they find themselves on a long stretch of single-lane highway with no passing opportunities. Each driver has a different "ideal" speed at which he would like to drive, and drives at that speed if unimpeded. Eventually the cars will be divided into one or more "cohorts," each cohort consisting of an unimpeded driver and his impeded followers. What is the probability that the cars will break into 4 distinct cohorts?

Hint: Assume you've solved the problem for any number of cohorts, in the case where there are seven cars. Now introduce an eighth car, faster than all the others. If it lands in the front position, it will add one new cohort (itself). Otherwise, it will nestle up against the car in front of it, adding no new cohort. Therefore, the chance that there will be four cohorts can be derived from the chances that the original seven cars generated either three or four cohorts.]

What if none of these approaches seem to work? Well, there's no simple answer to *that* question (and that's what keeps probability problems interesting). One fallback solution is to simulate the random process under investigation, and draw statistical inferences about the desired probabilities from the simulation results.