Uniform finite probability distributions
Math 217 Probability and Statistics

Prof. D. Joyce, Fall 2014

We've looked at a couple of these already, for instance, flipping a fair coin, or tossing a fair die. In general, a uniform distribution on a finite sample space Ω with n outcomes assigns to each outcome Ω the same value $1/n$. Therefore, the probability of an event E is the number of outcomes in E divided by n:

$$P(E) = \frac{|E|}{|\Omega|} = \frac{|E|}{n}.$$

That means for uniform finite probabilities, you can figure out the probabilities if you can count the outcomes in an event. Thus, everything we discussed about combinatorics applies directly to uniform finite probabilities.

Odds. The earliest studies of probabilities were all about this uniform finite situation, but they weren't expressed in terms of probabilities, that is, values between 0 and 1. Instead they were expressed in terms of 'odds.' These odds were usually not abstract numbers, but monetary payoffs for bets. Odds were the standard way to understand probability up through the 17$^{\text{th}}$ century.

Let's take a concrete example. Suppose we're tossing a pair of dice repeatedly until either a sum of 7 comes up or a sum of 5 comes up. On a single toss of a pair of dice, a sum of 7 occur up 6 different ways $(1+6, 2+5, 3+4, 4+3, 5+2,$ or $6+1)$, while a sum of 5 can occur in only 4 different ways $(1+4, 3+2, 2+3,$ or $4+1)$. In this game, there are 10 outcomes, which, assuming the dice are fair, all have the same probability. Since 6 are in the event $E = \{\text{sum is } 7\}$, while 4 are in its complement $E^{\mathsf{c}} = \{\text{sum is } 5\}$, we can conclude $P(E) = \frac{6}{10} = \frac{3}{5}$, while $P(E^{\mathsf{c}}) = \frac{4}{10} = \frac{2}{5}$.

We can express this result in terms of odds. Since E includes 6 outcomes, and E^{c} includes 4 outcomes, and each outcome is equally likely, we can say the odds of E to E^{c} are in the ratio 6 to 4, or, in lowest terms, 3 to 2.

Suppose you bet on E and your opponent bets on E^{c}. Then a fair bet is where you bet \$3, and your opponent bets \$2. If the game results in E, you collect all \$5, but if it results in E^{c}, your opponent collects all \$5. Thus, for you, \$3 will get you \$2, and for your opponent, \$2 will get your opponent \$3.

It's easier to deal with straight probabilities than with odds for two reasons. One is that probabilities are single numbers between 0 and 1 whereas odds are ratios of two numbers. The other is that probabilities don't involve money, wagers, and payments, but odds do, at least the historical origins of odds involve money, wagers, and payments.

On the other hand, probabilities are more abstract than odds since the concrete connection to real-world implications (gain or loss of money) is removed by one step.

Repeated trials. One of the most important questions for probability and statistics concerns repeated trials.

For example, if you toss 4 coins, what's the probability that 2 of them come up heads? For one coin, the sample space consists of two outcomes, $\{H, T\}$. For 4 coins, there are 16 outcomes in the sample space. For instance, the outcome $HTTT$ says the first coin comes up heads, the second tails, the third tails, and the fourth tails. Since there are 16 different strings of length 4 made out of H and T, therefore $|\Omega| = 16$. They should all have the same probability (since the outcome of one coin doesn't affect the outcome of any other), so we're dealing with uniform discrete probability. Each of the 16 outcomes has probability $\frac{1}{16}$. To determine the probability that 2 of them come up heads all we have to do is determine how many outcomes have exactly 2 heads and divide by 16. The number of ways to choose 2 to be heads out of the four coin tosses is the binomial coefficient $\binom{4}{2} = 6$. Thus, that probability is $\frac{6}{16} = \frac{3}{8}$.

In general, the probability of flipping exactly k heads in n flips is $\binom{n}{k}/2^n$.

We can do something similar for rolling dice. Suppose we want to know the probability of tossing k 5's when we toss n dice (or when we toss the same die n times). For one die, the sample space consists of 6 outcomes $\{1, 2, 3, 4, 5, 6\}$. For n dice there are 6^n possible outcomes. How many of these outcomes have exactly k 5's? It could be any k of the n tosses that come up 5's, and there are $\binom{n}{k}$ of those, but for the remaining $n-k$ tosses, they could come up any of the remaining five numbers 1, 2, 3, 4, or 6, and there are 5^{n-k} ways to do that. So altogether, there are $\binom{n}{k}5^{n-k}$ of them, each with the same probability of $1/6^n$. Therefore the probability is

$$\binom{n}{k}\frac{5^{n-k}}{6^n} = \binom{n}{k}p^k(1-p)^{n-k}$$

where $p = \frac{1}{6}$, the probability of getting a 5 on one roll.

This is an example of a binomial probability distribution.

Urns and balls. Lots of problems involving probability involve random choices of selection. Although they could be stated with different applications, their structure is similar and they're traditionally described as randomly selecting balls of various colors out of urns. It's always assumed that each ball that remains in the urn has the same probability of being chosen as any other.

Sampling without replacement. In this situation we have N balls in the urn, and M of them are "preferred", let's say black, and the rest, $N - M$, are white. You randomly select n of them, leaving $N - n$ of them in the urn. What's the probability of getting exactly k black balls (and $N - k$ white balls)?

Let's put in specific numbers to see it more concretely. Suppose there are $N = 11$ balls with $M = 4$ black ones and $N - M = 7$ white ones in the urn. You select $n = 5$ of them at random leaving the rest in the urn. What's the probability of getting exactly $k = 3$ black ones?

We can answer this with uniform discrete probabilities. We need to identify a sample space where each outcome is equally likely. There are 11 balls and we're choosing 5 of them, and any combination of the 11 is as likely as any other. Thus the sample space has $\binom{11}{5} = 462$ outcomes each with a probability of $1/462$.

How many of these combinations include 3 black ones and 2 white ones? The 3 black ones are chosen from the 4 black balls in the urn, and the 2 white balls are chosen from the 7 white urns, so there are $\binom{4}{3}\binom{7}{2} = 84$. Thus, the probability is $84/462 = 14/77 \approx 0.182$.

This is called *sampling without replacement* since we don't replace each ball after selecting it. The argument we just gave shows the the probability in general is

$$\frac{\binom{M}{k}\binom{N-M}{n-k}}{\binom{N}{n}}$$

when you have N balls, M of them preferred, select n of them, and get k preferred ones. These probabilities go by the name *hypergeometric probability distribution.*

Sampling with replacement. This is like sampling without replacement, but you replace each ball back in the urn before randomly selecting the next. That changes the probabilities.

We have N balls in the urn, and M of them are "preferred", let's say black, and the rest, $N-M$, are white. You select n of them, one at a time, replacing each one back in the urn before selecting the next. What's the probability of getting exactly k black balls (and $N-k$ white balls)?

Let's use the same specific numbers we had before. Suppose there are $N=11$ balls with $M=4$ black ones and $N-M=7$ white ones in the urn. You select $n=5$ of them with replacement. What's the probability of getting exactly $k=3$ black ones?

We can answer this with uniform discrete probabilities. We have a list of 5 balls we chose, each one can be any of the 11 balls in the urn, and they can be repeated since we're replacing them. So our sample space has 11^5 outcomes.

How many of those $11^5 = 161051$ outcomes have 3 black balls and 3 white balls? The 3 black balls can occur anywhere on the list of 5 balls, $\binom{5}{3} = 10$ ways of doing that. The other 2 places on the list will be for the 2 white balls. Those 3 places on the list for black balls can be filled by any of the 4 black balls in the urn, and there are $3^4 = 81$ ways of doing that; and the 2 places on the list for white balls can be filled by any of the 7 white balls in the urn, and there are $2^7 = 128$ ways of doing that. So, altogether, we have $30 \cdot 81 \cdot 128 = 31104$ of the outcomes that have 3 black balls and 2 white balls. Therefore the probability is $31104/161051 \approx 0.193$.

Generalizing that argument, you'll see that the probabilities in general for sampling with replacement are

$$\frac{\binom{n}{k} M^k (N-M)^{n-k}}{n^N} = \binom{n}{k} p^k (1-p)^{n-k}$$

where $p = M/N$ is the probability of getting a preferred ball when you choose just one ball out of the urn. This is another example of a *binomial probability distribution.*

The Birthday problem. What's the probability that among 23 randomly chosen people there are at least 2 with the same birthday? It's surprisingly large. To answer this question, we'll make a couple of simplifying assumptions. First, we'll assume there are only 365 possible birthdays. (You can add back Feb 29 to see how that affects the answer.) Second, we'll assume each is equally probable. This leads to a situation of uniform discrete probability. An outcome in the sample space consists of an assignment of birthdays to each of the 23 people. Since each of the 23 people can have any of 365 birthdays, there are 365^{23} outcomes in the sample space, each with the same probability.

Rather than find the probability that at least 2 have the same birthday, we'll compute the complementary probability, that is, the probability that all 23 have different birthdays. How many ways can that happen? What we want is a 23-permutation from a set of size 365. There are $365!/342!$ of them. Therefore, the probability that all 23 have different birthdays is $\dfrac{365!/342!}{365^{23}} = 0.4927028$, which is just under $\frac{1}{2}$. So the probability that at least 2 of them have the same birthday is just over $\frac{1}{2}$.

The Birthday applet computes these probabilities.

Math 217 Home Page at `http://math.clarku.edu/~djoyce/ma217/`