The Birthday/Lottery Ticket Problem

John D. Norton

There are N days in the year or N lottery ticket numbers available. We choose n days or n lottery ticket numbers, independently of each other, and with equal probability for each. What is the relationship between n, N and p, the probability that there are no duplications in the days or lottery tickets chosen?

The Exact Calculation

The probability that there are no duplications is given by

$$p = \frac{N}{N} \cdot \frac{N-1}{N} \cdot \frac{N-2}{N} \cdot \ \dots \ \cdot \frac{N-(n-1)}{N} = \frac{N!}{(N-n)!\, N^n} \tag{1}$$

Approximation with Stirling's Formula

We have from Stirling's formula that large factorials are well approximated as:

$$N! \approx \sqrt{2\pi N} \left(\frac{N}{e}\right)^N$$

Substituting into the expression for p we have

$$p \approx \frac{\sqrt{2\pi N}}{\sqrt{2\pi(N-n)}} \cdot \frac{(N/e)^N}{((N-n)/e)^{N-n}} \cdot \frac{1}{N^n}$$

The first term above simplifies to

$$\frac{\sqrt{2\pi N}}{\sqrt{2\pi(N-n)}} = \frac{1}{\sqrt{1-n/N}}$$

The second term simplifies to

$$\frac{(N/e)^N}{((N-n)/e)^{N-n}} = \frac{(N/e)^N}{((N-n)/e)^N} \cdot ((N-n)/e)^n = \frac{((N-n)/e)^n}{(1-n/N)^N}$$

The second and third terms together are

$$\frac{((N-n)/e)^n}{(1-n/N)^N} \cdot \frac{1}{N^n} = \frac{((1-n/N)/e)^n}{(1-n/N)^N} = \frac{(1-n/N)^n e^{-n}}{(1-n/N)^N}$$

Combining we have

$$p = \frac{1}{\sqrt{1-\frac{n}{N}}}\left[\frac{\left(1-\frac{n}{N}\right)^{\frac{n}{N}} e^{-\frac{n}{N}}}{1-\frac{n}{N}}\right]^N$$

(2)

Check formula

For $n = 23$ and $N = 365$, an exact calculation from (1) gives $p = 0.492703$. The formula (2) gives us $p = 0.492710$. The approximation is good to four significant figures.

Simplification for small n/N

Collecting terms up to second order in n/N, we have

$$\frac{1}{\sqrt{1-\frac{n}{N}}} \approx 1 + \left(\frac{1}{2}\right)\left(\frac{n}{N}\right)$$

Approximating $\left(1-\frac{n}{N}\right)^{\frac{n}{N}}$ is more complicated. Write

$$Y = \left(1-\frac{n}{N}\right)^{\frac{n}{N}}$$

Then we have

$$\ln Y = (n/N) \cdot \ln\,(1 - n/n) \approx (n/N) \cdot (-n/N) = -(n/N)^2$$

Recovering the expression from ln Y, we find

$$\left(1-\frac{n}{N}\right)^{\frac{n}{N}} = \exp(\ln Y) \approx \exp\,(-(n/N)^2) \approx 1 - (n/N)^2$$

We now have

$$\frac{\left(1-\frac{n}{N}\right)^{\frac{n}{N}} e^{-\frac{n}{N}}}{1-\frac{n}{N}} \approx \frac{\left(1-\left(\frac{n}{N}\right)^2\right)\left(1-\frac{n}{N}+\frac{1}{2}\left(\frac{n}{N}\right)^2\right)}{1-\frac{n}{N}} = \left(1+\frac{n}{N}\right)\left(1-\frac{n}{N}+\frac{1}{2}\left(\frac{n}{N}\right)^2\right)$$

$$= 1 - \frac{n}{N} + \frac{1}{2}\left(\frac{n}{N}\right)^2 + \left(\frac{n}{N}\right) - \left(\frac{n}{N}\right)^2 + \frac{1}{2}\left(\frac{n}{N}\right)^3 \approx 1 - \frac{1}{2}\left(\frac{n}{N}\right)^2$$

where the last approximation drops the third powers of n/N. Combining we find

$$p \approx \left(1+\left(\frac{1}{2}\right)\left(\frac{n}{N}\right)\right) \cdot \left(1-\frac{1}{2}\left(\frac{n}{N}\right)^2\right)^N$$

(3)

Approximation for n/N given p (for small n/N)

Inverting the approximation (3), we recover an expression for n/N. Raising (3) to the $1/N$ power, we have

$$1-\frac{1}{2}\left(\frac{n}{N}\right)^2=\left[\frac{p}{1+\left(\frac{1}{2}\right)\left(\frac{n}{N}\right)}\right]^{1/N}$$

Solving for n/N, we have

$$\frac{n}{N}=\sqrt{2\left[1-\left(\frac{p}{1+\left(\frac{1}{2}\right)\left(\frac{n}{N}\right)}\right)^{1/N}\right]} \tag{4}$$

Using (4) to compute n/N requires two steps, since the right-hand side of the equation also contains n/N. As long as the case is one of a small n/N, its value can be approximated by first computing n/N by assuming that n/N is zero in (4). That is, first compute

$$\frac{n}{N}=\sqrt{2[1-(p)^{1/N}]} \tag{5}$$

Then substitute the value recovered in (5) into (4) and use (4) to make the corresponding small adjustment to the value of n/N.

Even Simpler Approximation for n/N

If we approximate

$$1+\left(\frac{1}{2}\right)\left(\frac{n}{N}\right)\approx 1$$

we can recover a still simpler approximation for n/N. The approximation depends on taking a power series expansion in x for

$$f(x)=f(1/N)=1-(p)^{\frac{1}{N}}=1-p^x$$

where we set $x=1/N$. We need the first derivative of $f(x)$:

$$\frac{df(x)}{dx}=\frac{d}{dx}(1-p^x)=\frac{d}{dx}(-p^x)=-\frac{d}{dx}\exp(\log p\cdot x)=-\log\ p\cdot p^x$$

We form the power series expansion about $x=0$, which is equivalent to $N=\infty$.

$$f(x) = f(0) + x\frac{df(0)}{dx} + \cdots = f(0) - x \cdot \log p . p^0 + \cdots = -x . \log p + \cdots$$

since p^0 and $f(0) = 0$. Recalling that $x = 1/N$, we recover an approximation to first order in $1/N$:

$$(1 - p^x) \approx -(\log p)/N$$

Substituting this approximation into (5), we recover[1]

$$\frac{n}{N} = \sqrt{-2\,(\log p)/N} \tag{6}$$

This last formula (6) gives a rough picture of how n/N grows with increasing N, when p has a fixed value:

$$n/N \propto 1/\sqrt{N}$$

$$n \propto \sqrt{N}$$

Thus, after N is large, as N increases by a factor of 10 through 1000, 10,000, 100,000 etc., n/N *decreased* by a factor $\sqrt{10} = 3.16$ and n itself *increases* by a factor $\sqrt{10} = 3.16$. It follows that n can grow arbitrarily large with increasing N, but n/N will decreased arbitrarily close to 0.

[1] Square root of a negative number? No. Since $p < 1$, $\log p < 0$, so $-\log p > 0$.