

# What is the probability of winning the Hamilton lottery?

## Simple Statistics can tell you the answer

Dana Mastropole and Tianhui Li

January 28, 2017

People interested in seeing the Broadway musical [Hamilton](#) – and there are still many of them, with demand driving starting ticket prices to \$600 – can enter Broadway Direct’s daily lottery. Winners can receive up to 2 tickets (out of 21 available tickets) for a total of \$10.

### What’s the probability of winning?

How easy is it to win these coveted tickets? Members of NYC’s [Data Incubator Team](#) have collectively tried and failed 120 times. Given our data, we cannot simply divide the number of successes by the number of trials to calculate our chances of winning – we would get zero (and the odds, which apparently small, are clearly non-zero).

This kind of situation often comes up under many guises in business and big data. And because we are a [data science corporate training company](#), we decided to use statistics to determine the answer. If you’re measuring the click-through-rate of a piece of organic or paid content. Out of 100 impressions, you have not observed any clicks. The measured CTR is zero but the *true* CTR is likely not zero? Alternatively, suppose you’re measuring the rate of severe adverse side effects of a new drug. You’ve tested 40 patients and haven’t found any, but you know the chance is unlikely to be zero. So how great is it?

Without any positive cases, we cannot get a lower bound (lowest reasonable probability) on the probability beyond 0 since we have never seen any positive events. Fortunately, the [Rule of Three](#) gives us a reasonable estimate for the upper bound (highest reasonable probability) on the probability of success, given a certain number of failures. It states that with 95% confidence, the probability of success is less than  $3/N$ , where  $N$  is the number of failures. Given  $N = 120$ , we conclude that a reasonable estimate for the probability of winning the *Hamilton* lottery is **less than 2.5%**.

### How big is the competition?

How many people enter the *Hamilton* lottery? We know that winners receive either 1 or 2 tickets out of an available 21. Therefore, the fewest number of winners per lottery is 11 (10 two-ticket winners and 1 one-ticket winner). With this information, we can estimate a lower bound on the number of people entering the lottery. In order to get the minimum of 11 winners with a maximum success rate of 0.025, we conclude that **at least 440 people** enter daily. Of course, this is likely to be a very conservative underestimate.

### How do our numbers compare with the reported values?

According to a [Mashable Article](#) that came out earlier this year, **over 10,000 people** enter the lottery daily. If every winner is awarded a single ticket (i.e. there are 21 winners), the probability of winning is 0.21%. In the case where there are only 11 winners, the probability of winning is 0.11%. Therefore, the probability of winning the lottery is **between 0.11% and 0.21%**, less than one tenth of what we estimated. Of course, this is entirely possible – since we haven’t won a single ticket and could only calculate an upper bound, not a lower bound.

### How long do you have to wait to win?

The distribution of the number of failures up to and including the first success is given by the [geometric distribution](#), where the expected (or average) number of trials is  $1/p$ . Plugging in the high (0.0021) and low (0.0011) estimates for  $p$ , we expect to enter the lottery between 476 and 909 times:

Assuming that *Hamilton* does not continue to run on it’s current schedule for the next 2-3 years, how long do we expect to wait for 476 to 909 performances to occur? We can get a very rough estimate for the number of performances per day by considering some of the [longest running Broadway shows](#).

Show	Run Time (Days)	Performances	Avg. Daily Performances
The Phantom of the Opera	10,591	12,062	1.14
Chicago	7,386	8,387	1.14
The Lion King	7,012	7,986	1.14
Cats	6,548	7,485	1.14
Les Misérables	5,911	6,680	1.13

If we divide the expected number of performances (476 and 909) by the average daily performances (1.14), we get a wait time of **340 to 649 days** (one to two years!).

**Derivation of the Rule of Three:** Let’s say that the probability of winning the lottery is  $p$ . We know from probability theory and statistics that the discrete probability distribution of  $k$  successes in  $n$  trials is binomial.

$$\Pr(X = k) = \binom{n}{k} p^k (1 - p)^{n-k}$$

It follows that the cumulative distribution function is:

$$\Pr(X \leq k) = \sum_{i=0}^k \binom{n}{i} p^i (1 - p)^{n-i}$$

In the case where we have zero successes ( $k = 0$ ), the right hand side of the equation reduces to  $(1 - p)^n$ . If we want to calculate the  $(1 - \alpha)\%$  confidence interval for the upper bound of  $p$ , we then set the right hand side of the equation to the significance level,  $\alpha$ .

$$\begin{aligned}(1 - p_{upper})^n &= \alpha \\ p_{upper} &= 1 - \alpha^{1/n}\end{aligned}$$

Doing a Taylor Series expansion for  $\alpha^{1/n}$ , we get:

$$\alpha^{1/n} = 1 + \frac{\ln(\alpha)}{n} + \frac{[\ln(\alpha)]^2}{2n^2} + \dots$$

Dropping the non-linear terms and substituting in for our equation for  $p_{upper}$ , we get:

$$p_{upper} \approx -\frac{\ln(\alpha)}{n}$$

For  $\alpha = 0.05$ , and  $-\ln(\alpha) \approx 2.9957$  we get:

$$p_{upper} \approx \frac{3}{n}$$

We just derived the Rule of Three! Plugging in our value for  $n$ , we get  $p_{upper} \approx 0.025$ .