Measurement of Binomial Failure Rates

Benjamin Lucas - February 8, 2018

Testing seeks to measure how frequently some event happens. This could be the rate of system failures in a launch routine, the rate of user errors in completing a task, or the rate of photon detection in an radioactive decay source. In just about every single job I have had, measuring the rate of failure/success has come up. Most of these scenarios follow a binomial distribution. While the math behind this is fairly simple the consequences of this math seems to be very counterintuitive - to the extent that people with advanced degrees in quantitative fields (Physics, Chemistry, Electrical Engineering, Computer Science…) often overlook these effects. In the interest of better understanding I have written up a basic exposition of how to calculate error rates and their uncertainty in basic binomial situations.

For a pass/fail measurement that is sampled with replacement (each trial is uncorrelated with past and future events) the distribution of failures/successes follows a Binomial Distribution.

When the number of samples (*N*) is large and the probability of failure (*p*) is not close to 0, or 1, the binomial distribution can be approximated as a normal distribution. As a rule of thumb, this approximation works when $N \cdot p > 10$.

In this scenario, error of the failure rate measurement can be calculated on the basis of the number of measurements N and the measured failure rate $p_{meas} = \frac{n_{fail}}{n_{total}}$ from the following simple equation:

$$error = z \cdot \sqrt{\frac{p_{meas} \cdot (1 - p_{meas})}{N}} \quad \text{(Eq 1)}$$

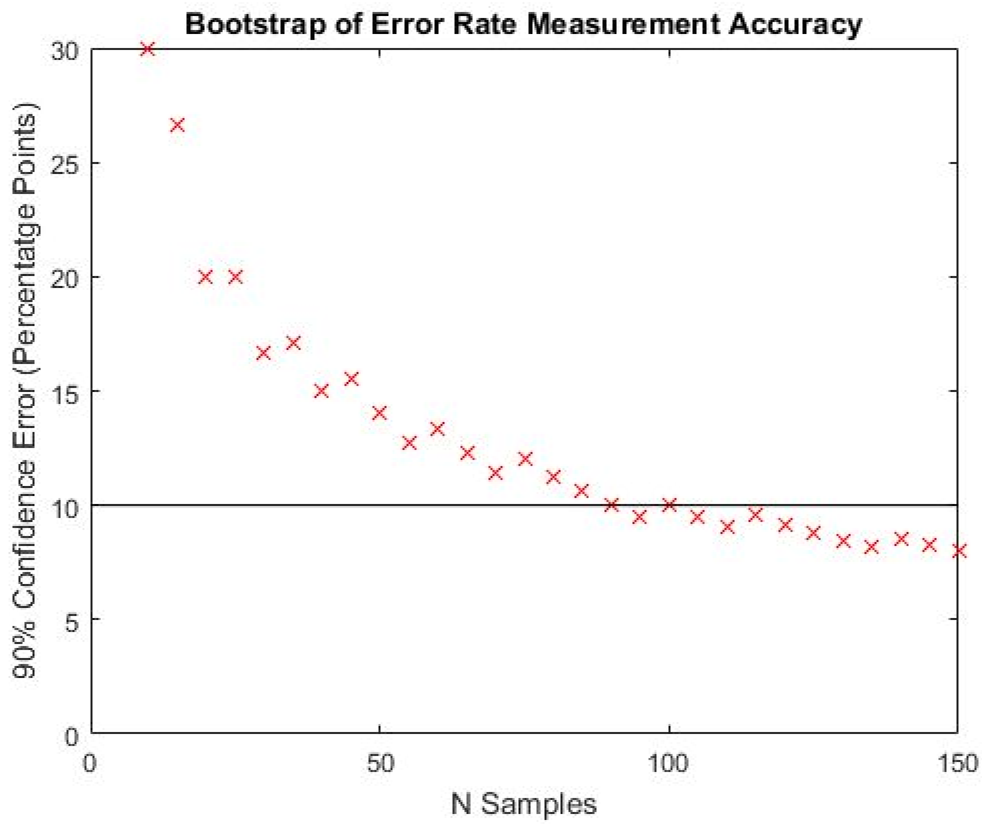
Here *z* is a constant that relates the cumulative probability density of the normal distribution and the required confidence level. For a 95% confidence measurement z = 1.96; for a 90% confidence measurement z = 1.64.

Let's say we make 20 measurements and we see that 3 result in failures - so we estimate the failure rate as being ~15%. How confident are we in this rate? The binomial distribution is not a great approximation here ($N \cdot p \approx 3 < 10$), but for the sake of argument, let's just plug into equation 1. Using this equation with a 90% confidence interval, we see that the error in the estimate is ~13%. In other words if we made the measurement again, we should not be surprised to see anywhere between 0 and 6 errors in 20. The error of the measurement is as big as the signal we are trying to measure!

So how many measurements do we need? It depends on the expected error rate, and the accuracy that we need on this measurement. If we assume that the failure rate is around 10%, and that we would like to have a 5% accuracy on this measurement with 90% confidence (p = 0.1 +/- 0.05), then equation 1 predicts that we will need to make **97 measurements**.

The normal approximation of the binomial distribution is great - it gives us a convenient equation to calculate results, but what if we wanted to use the precise distribution? Another way to look this question is to use a bootstrap method. Here the outcome of several trials is simulated, and the size of the confidence interval is measured directly. This avoids the need for any approximations.

Using this method we estimate the accuracy of our measurement computationally. If we expect a failure rate p = 0.1, and we would like to have +/- 0.05 on the accuracy of this measurement (the full confidence interval is 0.1) then we see the following result:



The above figure confirms that measuring the failure rate that is around p = 0.1 to within +/- 0.05 to 90% confidence requires on the order of 100 measurements.

Note that the requirements of this will go up more or less quadratically with the requirement for precision. Thus if we would like to be able to state the the failure rate if 0.1 +/- 0.025 then we will need on the order of 400 measurements. For a quick look up table, see the appendix.

Appendix:

Table of Accuracy:

Expected Error Rate:	10% Tolerance:	5% Tolerance:	1% Tolerance
50%	96 Trials	384 Trials	9604 Trials
40%	92 Trials	369 Trials	9218 Trials
30%	81 Trials	323 Trials	6067 Trials
20%	61 Trials	246 Trials	6144 Trials
15%	48 Trials	196 Trials	4898 Trials
10%	34 Trials	138 Trials	3457 Trials
7%	25 Trials	100 Trials	2501 Trials
5%	18 Trials	73 Trials	1825 Trials
3%	11 Trials	45 Trials	1118 Trials
2%	8 Trials	30 Trials	753 Trials
1%	4 Trials	15 Trials	380 Trials