Starting Strength

Mathematical Modeling of Attempt Selection

by

Jeff Russell

One Saturday afternoon you decide to head to the gym and max out your squat. Maybe it's been a while, and you want to see if you can set a new personal record - a number you've had your eye on. Once you arrive, you put on your shoes and begin to warm up the lift. After a few sets and some singles, you have your PR attempt loaded on the bar in front of you. You steel yourself, get under the bar, walk it out, squat down and… miss the attempt. Shit.

Being a responsible gym-goer, you had two buddies on hand to bail you out. So you're still standing and you haven't been kicked out for damaging the equipment, but you now have a decision to make: will you re-attempt the weight? The first try felt close; you almost made it out of the hole. But something wasn't quite right. Aside from feeling heavy, it didn't seem like a normal squat.

After a few minutes of deliberation you decide to try again. Reciting cues in your head and doubling your resolve, you approach the bar, walk it out, squat down and… pop back up. Easy! At least relatively so. Checking the video, you see the depth was good. Your buddies are duly awestruck by your performance. You are now an Instagram celebrity, and you all head out for cheeseburgers to celebrate.

Reliability

The results of near-limit attempts like this are informative in multiple ways. First, you can determine that you are unambiguously strong enough to move a given weight. This can be useful for planning training, picking attempts in competition, or just personal satisfaction. But there is information in any misses as well. In the above "new PR" scenario, you have just measured two things, numerically speaking:

1. Your physical limit (1RM) is *at least* this weight.

2. You've made *one of two* attempts at this weight.

Misses can occur for innumerable reasons. Maybe you lost tension at the bottom, didn't set your back correctly, the bar slipped, you got off balance, you inhaled a fly, started blacking out, et cetera. Insufficient strength will certainly cause a missed attempt, but errors in execution can and often do cause misses at achievable weights. In addition, many lifters swear they have "off days", where some fraction of strength is inexplicably, temporarily unavailable. The grim spectre of injury is also ever-present. In short, variability in performance abounds, even under controlled circumstances.

A more formalized notion of the *reliability* of a lift may be useful. Here we define reliability as the chance of success of a given lift at a given weight. In statistical terms, it is the probability of success (**p**), a numerical value between zero and one. This value is fairly intuitive. You may have heard lifters mention certain weights as being "reliable", meaning an attempt at that weight is very likely to succeed (**p** is close to 1). Similarly, an "unreliable" weight, most likely a heavy near-limit attempt, will be much less likely to succeed (**p** is closer to 0).

For any given lifter, the reliability of a lift will vary based on the weight attempted, with a rather nonlinear and precipitous drop right under the lifter's strength limit. It may look something like this:
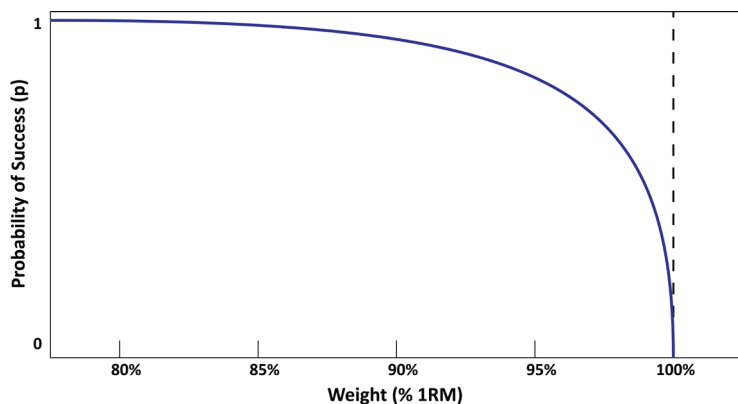
The shape of this curve will vary based on several factors, but is generally indicative of a lifter's *proficiency* in a lift. A proficient lifter will have a steeper "cliff" in the curve, and thus higher odds of success at relatively heavy weights. Conversely, an unproficient lifter will see an earlier and more pronounced degradation in reliability as weight increases:
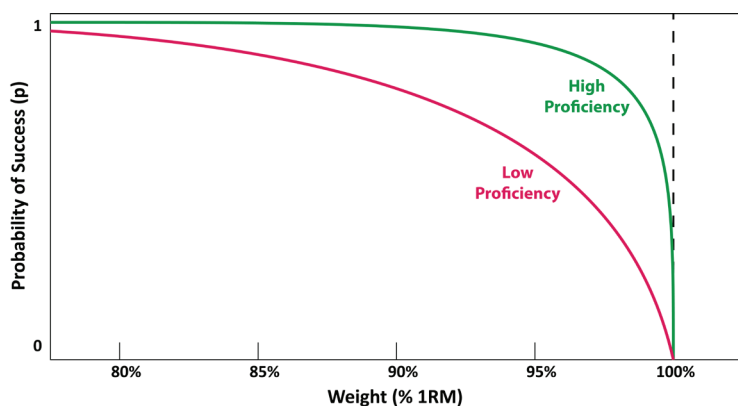
Most novices are unproficient lifters. They may be lacking in technique, consistency, and muscle recruitment to the point where they are unlikely to perform a near-maximal lift at all. A clumsy person or someone with glaring form issues may have similarly low proficiency. An intermediate lifter should exhibit much more proficiency and hence reliability with heavy weights. Advanced lifters and elite athletes may have very high proficiency and as a result very "steep" reliability curves.

Different lifts may also exhibit different reliability for a given athlete. A deadlift, for example, is relatively straightforward for most, whereas a squat may pose more technical problems and a snatch might be trickier still. Each event will have its own profile, depending on the strength, skill, and general efficiency of the lifter.

Measurement

In our hypothetical above, you had made one of two attempts at a heavy weight. This "computes" to a reliability of 0.5 - a 50/50 shot. The problem of course is that data in this situation are desperately sparse, and two measurements are insufficient to be confident in this value. A third attempt might succeed, yielding a new probability of 2/3 (**p** = 0.67) - or fail, giving a new probability of 1/3 (**p** = 0.33). Fourth and fifth attempts may change the estimate significantly as well. Measuring experimental probability requires many trials before we can have any confidence in our estimate.

 StartingStrength.com

Unfortunately, this is difficult to do in a controlled way. You could decide to go do twenty heavy singles at varying weights tomorrow, but fatigue is going to interfere with your results significantly even if you manage to finish. To better manage the fatigue you could spread the trials out over a few weeks, but if your training is working at all in the meantime you'll be getting stronger, which means we're now measuring a moving target. And even if fatigue and changing strength could be set aside, there is one more problem: doing lots of heavy singles tends to make you better with them. All the practice will improve your proficiency, an otherwise welcome development but a confounding factor in measuring your proficiency today.

So we are left with a quantity which is difficult to measure and tends to change around on us when we try. Nevertheless, most of us are engaged in constant estimation of our proficiency. We extrapolate a kind of mental curve, albeit in a sort of fuzzy human way, based on recent singles and sets performed with each lift. We use this impression to select attempts for competition or gym PRs. The human brain is actually fairly good at finding patterns in sparse data (it seems likely we evolved to do this), and so an experienced coach or athlete can often pick attempts with surprising accuracy.

Lifters also have access to a kind of internal feedback beyond a simple make/miss notation; they can judge a lift by how it felt and what may have gone wrong. Maybe a lift went up despite major errors, in which case we judge it to be submaximal. Or perhaps a near miss occurred with an error present, which might cause us to file that weight as achievable with better execution. Our experience, even if limited, is full of variables we can correlate to make our final estimates.

Despite the "fuzzy" nature of this process and a possible dearth of hard data, it is this author's intent to show that some general lessons can be learned from the strictly mathematical aspects of a lift's reliability. With imperfect or even hypothetical data, useful patterns emerge that may enhance our understanding of this phenomenon. If nothing else, the value of thinking of the matter in statistical terms, rather than absolute limits, may become clear.

Competition & Selection

Organized competition provides a good framework for examining the shape and meaning of a lift's reliability. The rules for competitions vary, but for powerlifting and weightlifting there are some common basics: each lifter gets three attempts in each event, and a lifter may repeat or increase a given weight but may never drop the weight to a lower value than a previous attempt. In this way lifters make multiple attempts and the highest successful lift is the competitor's score for a given event (or zero if no attempts are successful).

Within the framework of these rules, a competitor must select weights with an optimal balance of risk and reward. Light weights are likely to succeed, but may not place an athlete well in the competition. Heavy weights may place well if successful, but may be likely to fail. Selecting the perfectly correct balance is difficult. Ideally a lifter will make all attempts without selecting weights that are too easy.

With some knowledge of a lift's reliability over different weights (that is, a reliability curve) a simple algorithm can select the three attempts that maximize the *average* successful attempt. This may not necessarily be the lifter's goal (more on this later), but aiming for the highest average result will give a pretty good guideline for attempt selection in the general case of a lifter wanting to do well in a meet. Our goal is to find three weights, $\mathbf{w}_1$, $\mathbf{w}_2$, and $\mathbf{w}_3$, that yield an optimal outcome. A lifter will take $\mathbf{w}_1$ as the first attempt, and then if successful, take $\mathbf{w}_2$ as the second attempt, and so on. We'll keep things simple and assume the lifter will repeat any missed weights on successive attempts.

 StartingStrength.com

With our hypothetical reliability curve, we can determine the probability of a successful attempt of each weight, which we will correspondingly call $\mathbf{p}_1$, $\mathbf{p}_2$ and $\mathbf{p}_3$. Each $\mathbf{p}_n$ is determined by evaluating the reliability function, $f()$, with a given weight $\mathbf{w}_n$:

$$\mathbf{p}_n = f(\mathbf{w}_n)$$

A single meet event consisting of three attempts has four possible outcomes: either the lifter misses all attempts (score of 0), or only makes their opening weight (score of $\mathbf{w}_1$), or makes their first and second (score of $\mathbf{w}_2$), or makes all three attempts (score of $\mathbf{w}_3$). We can determine the probabilities of these outcomes by chaining our $\mathbf{p}$ values together for each series of misses and makes that would produce a given outcome, with probability of success for a weight being $\mathbf{p}_n$, and probability of a miss being $1\text{-}\mathbf{p}_n$. If the math here seems a little inscrutable, hang around and we'll see the same thing in plainer terms in a moment:

$\mathbf{P}_0 = (1\text{-}\mathbf{p}_1)^3$	*Lifter "bombs out"*
$\mathbf{P}_1 = (1\text{-}\mathbf{p}_1)^2\bullet\mathbf{p}_1 + (1\text{-}\mathbf{p}_1)\bullet\mathbf{p}_1\bullet(1\text{-}\mathbf{p}_2) + \mathbf{p}_1\bullet(1\text{-}\mathbf{p}_2)^2$	*Lifter makes only first weight*
$\mathbf{P}_2 = \mathbf{p}_1\bullet(1\text{-}\mathbf{p}_2)\bullet\mathbf{p}_2 + (1\text{-}\mathbf{p}_1)\bullet\mathbf{p}_1\bullet\mathbf{p}_2 + \mathbf{p}_1\bullet\mathbf{p}_2\bullet(1\text{-}\mathbf{p}_3)$	*Lifter makes only first and second weights*
$\mathbf{P}_3 = \mathbf{p}_1\bullet\mathbf{p}_2\bullet\mathbf{p}_3$	*Lifter makes first, second, and third weights*

The average score of a meet with the above outcome probabilities is then the sum of these probabilities multiplied by their corresponding weights:

$$\mathbf{R}_{avg} = \mathbf{P}_3\bullet\mathbf{w}_3 + \mathbf{P}_2\bullet\mathbf{w}_2 + \mathbf{P}_1\bullet\mathbf{w}_1 + \mathbf{P}_0\bullet 0$$

As stated above, our goal is to maximize the average meet result ($\mathbf{R}_{avg}$) by selecting optimal weights ($\mathbf{w}_1, \mathbf{w}_2, \mathbf{w}_3$). Maximizing a nonlinear function with this many variables can be a challenging task, but fortunately our range of feasible values (weights that humans can lift) is small enough that a simple exhaustive search can be made with software.

To make this a bit clearer, let's consider and compute an example. Say we have a lifter who needs to select attempts for the squat. This lifter feels that anything less than about 485 lbs is basically a sure thing, but anything over about 535 lbs is pretty much impossible. This means that the interesting part of the reliability curve, where the odds of success drop off towards zero, will lie between 485 and 535. We can create a reliability curve in this range, and then run our above analysis to determine three attempts with the highest average outcome. The results in this case look like this:
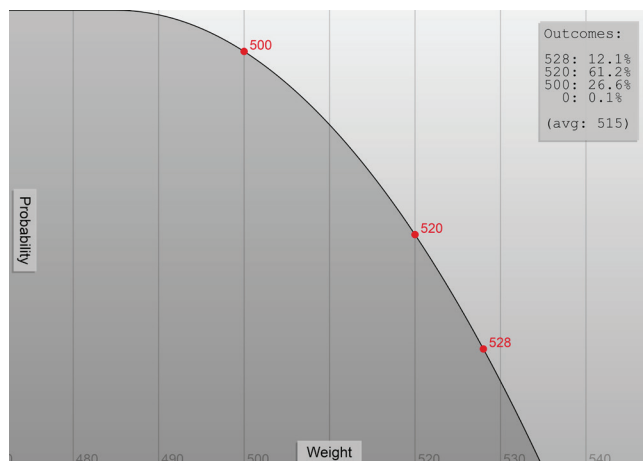
This process has selected three weights to be taken as three attempts: 500, 520 and 528. The average result is approximately 515, which represents a maximum of all possible weights in this range along the reliability curve. Probabilities for each of the four outcomes, including missing all attempts, are listed in the upper right.

It's important to reiterate that the above curve, the line drawn on the graph, is *hypothetical* and not based on much real data

in this case. It is used here to provide a plausibly smooth transition between the sure-thing weight and the impossible weight given by the lifter. The three labeled attempts are chosen to produce the highest average score based on this curve. If better data exist, as in the case of a lifter who has tried and recorded many singles in this range, then empirical data could theoretically be used in place of this curve in the same fashion.

The software used to produce the above results can produce plots and values for any sensible range. It is linked here; try it out for a little bit with some of your own numbers and observe the results. The "Min" and "Max" fields should be set in a fashion similar to our above example - with the "Min" weight being the heaviest weight that seems like a sure thing, and the "Max" weight being the lightest weight that seems too heavy to lift.

[Try the program - runs in the browser; try it!]

Note also the "Curve" setting, and the effect its change has on the optimal weights. This setting selects from a few different curve shapes, without modifying the given range of weights. As you may recall, the shape of the curve in our model is closely related to the proficiency of the lifter, with "steeper" curves representing more proficiency. The "Polynomial" options here represent higher proficiency than the simple "Linear" curve. The "Cumulative Distribution" curve is included as a kind of hybrid option, and because it likely to be applicable here for statistical reasons.

Results & Observations

Looking at these attempt selections, with a variety of reliability curves, a few trends begin to emerge. It's perhaps worth mentioning that the software selection was not groomed to produce any particular pattern: these results are simply products of selecting for the highest average total, given odds of success for various weights. A lot of conventional wisdom about attempt selection for meets seems to be reinforced to a large degree, even with this rather abstract model.

One of the first trends to stand out is the relative distribution of the three selected weights; these values all seem to have similar relative distribution almost regardless of the shape of the curve. Specifically, the second and third weights tend to be significantly closer to one another than to the first. This seems to occur with curves of most any shape, in situations of both high and low proficiency. This shouldn't be a big surprise to many competitors, who are used to taking a safe opener and then more aggressive second and third attempts, with the jump from second to third often being smaller than from first to second.

This pattern suggests that an even spacing of attempts, where the difference between first and second is similar to the difference between second and third, is likely to be suboptimal in terms of the aggregate result. Even worse would be taking a bigger jump from second to third weights than from first to second, likely meaning the first two attempts were too light, or the third ludicrously heavy. **The sweet spot seems to lie in making a smaller jump from the second to third attempt**. If a competition consisted of more than three attempts, we would likely see the same pattern continue: successive attempts coming at diminishing increments to balance the risk/reward factor in light of the increasing odds of failure.

Another interesting but perhaps unsurprising result is the selection of the opening attempt as a pretty safe value (generally around 90% odds of success). It is worth noting that the algorithm is given a range of weights with higher probabilities of success at essentially 100%, but it seems to always forego them in favor of a heavier and slightly riskier opener. **A safe but not entirely risk-free opener seems to be optimal.** This makes a bit of sense at an intuitive level: a heavier opener gives "more room" to take

riskier second and third attempts, since a relatively high number has already been posted. Additionally, a lifter has as many as three attempts to lift the opening weight, so accepting some modest risk with the opener does not impose a significant hazard of "bombing out" and making no total.

Another trend presents itself with steeper reliability curves (that is, lifters with higher proficiency). In all cases attempts tend to bracket the steepest portion of the curve, and making the curve steeper brings the attempts closer to one another. (Try it: switch back and forth between the "linear" and "polynomial (steep)" curve modes in the program.) **This may imply that lifters with higher proficiency can and should open at higher fractions of their limit**, as these weights carry less risk for them than they would for others.

Lastly it should be noted that the attempt selections made here result in some rather aggressive second and third attempts. Second weights are chosen at around 50% odds, and thirds end up even lower at around 20-30% odds. **This means the chances of making all three attempts are generally less than 20%, making it very likely a lifter will miss one or more attempts with these selections**. This degree of risk may seem high, but is not at all unheard of in very competitive circumstances (if you watched 2018's World Weightlifting Championship, for example, you probably saw some A-group sessions with more red than white on the board). Our algorithmic selection is optimizing for highest average total, and makes no attempt to avoid misses except insofar as they affect this average. Put another way: if you optimize for your total, you'll get more total, at the cost of other considerations.

Other Goals & Factors

It's important to understand that there are several limitations to this kind of simple model. Analyzing an idealized curve in a user-defined range, as we have done above, not only makes use of somewhat incomplete data but fails to account for other complicating factors in real competitions. In addition, optimizing for the average total may not be what every lifter wants to do; goals often differ.

Under very competitive circumstances, where one athlete may be vying with another or trying to reach a specific placement, probabilities and aggregate statistics become much less important. Similar exceptions apply in cases where an athlete wants to PR a certain lift, or achieve some other fixed goal. In these situations, attempts should be selected with their accordingly specific aims in mind and with much less emphasis on the average outcome.

Some athletes, in particular those new to competition, may find the degree of risk outlined above to be a problem. Missing lifts in competition can be mentally taxing even to experienced competitors, and new lifters may get into trouble dealing with missed openers and the like. Proper coaching and mental preparation is key of course, but in some cases backing off from the mathematically prescribed razor's edge of risk may be wise. Going into competition with a plan that has you very likely to miss some attempts that day requires a degree of mental fortitude and level-headed thinking that may take time to develop.

Another major complicating factor is that of fatigue. After a lifter has attempted several heavy lifts, the probability of success for additional heavy attempts will usually shrink. Meets consisting of multiple events may find the athlete in a diminished state by the last event, and attempt selection should adjust accordingly. Final deadlift attempts at the end of a full power meet are notably different events than setting a deadlift PR in the gym, for example. This is less of a factor in weightlifting, where the weights in the snatch and clean and jerk are lighter and fatigue less of a detriment.

On top of this, the model does not account for possible correlation in the outcome of attempts. If you enter a meet after a recent illness, for example, you may be more likely to miss all of your attempts.

In situations like this, re-attempting the same weight after a miss may not bring very good odds of success. Additionally there is a psychological component at work in some athletes after a miss which can make the following attempt more difficult than it might otherwise be. These factors are difficult to account for analytically, and so the model simply assumes that each attempt is a probabilistically independent event.

Lastly, our model does not consider the strategic possibility of increasing weight after a failed attempt. This is sometimes done to remain competitive in cases where a missed attempt is seen to be easily correctable. Whether and when to do this is something a mathematical model probably can't suggest (the software selection currently assumes a lifter repeats a missed weight), but it can perhaps inform the choice of increment in the event of a miss. From a mathematical perspective, a missed opener is a bit like partaking in a meet with only two attempts: you'll need to increase your "opener" (now second attempt) a little bit to account for the lack of following attempts. Some preliminary experiments with this have shown only modest value in increasing weight after a miss, provided the lifter is already using the rather aggressive attempts suggested by the model.

For the above reasons, and likely others unstated here, automated recommendations as produced by models such as this one should augment but never replace human judgement. The software used above is not really meant as an "attempt selector" so much as a tool for studying probability in competition. Goals and human factors vary enough that caution should be used to avoid placing too much faith in these results. They are, after all, only as good as the data and assumptions fed into them.

Conclusion

With a basic mathematical model for attempt success, basic strategies for attempt selection present themselves fairly clearly. Perhaps surprisingly, winning strategies in the model closely match winning strategies (or at least, common wisdom) in real world competition. Mathematical modeling of this type is valuable for developing an analytical understanding of risk and reward, as well as general thought on the subject of heavy singles and competition. Hopefully your own thinking has been honed to some degree with these exercises. Attempt selection is a difficult problem to solve fully, but with more tools at our disposal we may all improve. Choose wisely.

Starting Strength : Articles : Videos : Podcasts : Training : Forums