

External Oracle Price Updating: When is Enough Enough?

0xULT

0xULT@protonmail.com

October 13, 2021

Abstract

As trading volumes within the DeFi ecosystem continue to grow at an astounding clip, Automated Market Makers (AMMs) have been finding new and creative ways to reduce the impact of impermanent loss. One of the recent trends that several DEXes have started to adopt is the integration of an external price oracle to allow for price discovery while allowing their liquidity providers to skip some of the pain that comes from being exposed to informed trading. Within this paper we will explore an option around how to optimize the updating of oracle prices to find a good balance between smoothness of updates and minimization of arbitrage opportunities.

Usage of Oracles By DEXes

The issue of impermanent loss is pervasive in the economics of cryptocurrency markets. Historically the design of AMMs have forced liquidity providers to accept the fact that their funds will be traded against by more informed traders while hoping to more than offset the losses through a per swap fee. Impermanent loss is in essence the lost opportunity cost of engaging in AMM and having your portfolio be rebalanced in such a manner by external traders that you are worse off versus if you just held your funds instead. Several decentralized exchanges have started to implement new AMMs that leverage an external price oracle to periodically change the pool's internal price to match the external market's global price [[1], [2]]. This approach has the benefit of theoretically allowing for less dependency on arbitrageurs to aid in price discovery and push the internal pool price to match the external market's price. Instead the usage of the external oracle will act to some degree as the main price discovery mechanism. The downside of this approach though is that there are negatives associated with depending upon an external oracle: 1) a non zero transaction cost exists for updating the AMM's smart contract and 2) blindly depending upon an oracle may introduce additional noise into the internal pool's pricing curve. Throughout this paper we will develop a proposed solution that could help protocol developers more effectively integrate external market prices into their AMM.

Current Pros and Cons of Oracle Usage

A common approach for AMMs that leverage an oracle for price updating is to allow an update when there is a certain percent change in the broader market price. The positives of taking this approach is that it is simple to implement and predictable around what qualifies as a need to update or not. A downfall with it though is that in times of high volatility the price may move more quickly then price updates through the oracle can realistically occur. Additionally, single point thresholds may introduce additional noise into the AMMs price feed as it could just be reacting to noise more often then not. Given this uncertainty we hope that the proposal of this paper offers an alternative to the standard fixed percent change approach.

Optimal Updating Time

Historically integration of oracles have focused on what sort of price changes should trigger an oracle price update. This perspective can actually be shifted to find that it is analogous to instead look at oracle updates from the viewpoint of how often should an AMM be querying the oracle. To understand the optimal frequency with which we should be matching the internal price of an AMM to the market, we need to first establish a simple model to measure the loss (L) of incorporating an external oracle into a hypothetical AMM:

$$\begin{aligned} E[L] = & \alpha(P(X \geq 0) \min\{0, UB - E[X|X \geq 0]\} + \\ & P(X < 0) \min\{0, E[X|X < 0] - LB\}) \frac{S}{t} - \frac{G}{t} \end{aligned} \quad (1)$$

With the following notation and assumptions:

- α represents the prevalence of arbitrageurs in the market. Is a probability measure bounded from 0 to 1
- X is the price change of the external exchange rate between the two tokens in a pool over the time period t . X follows a lognormal distribution [6] with annualized price change of μ and standard deviation of annual returns σ . Per period mean is μt and standard deviation is $\sigma\sqrt{t}$
- $P(x)$ denoting the probability of x occurring
- LB and UB are respectively the no arbitrage lower bound and upper bound. These bounds contain the LP's current swap price within it and as long as the external market's price is within the range then arbitrageurs cannot profitability trade against the DEX. More in detail about these components below
- S is a scaling parameter that translates the change in price to a USD denominated change and is dependent on the size of the pool
- G signifies the average gas cost to receive an update from the oracle in USD
- t represents the time interval wait between when we are querying the price oracle and is expressed as percent of a year (i.e. $t = 0.5$ means we wait 6 months in between updates). $\frac{1}{t}$ gives us the number of updates per year

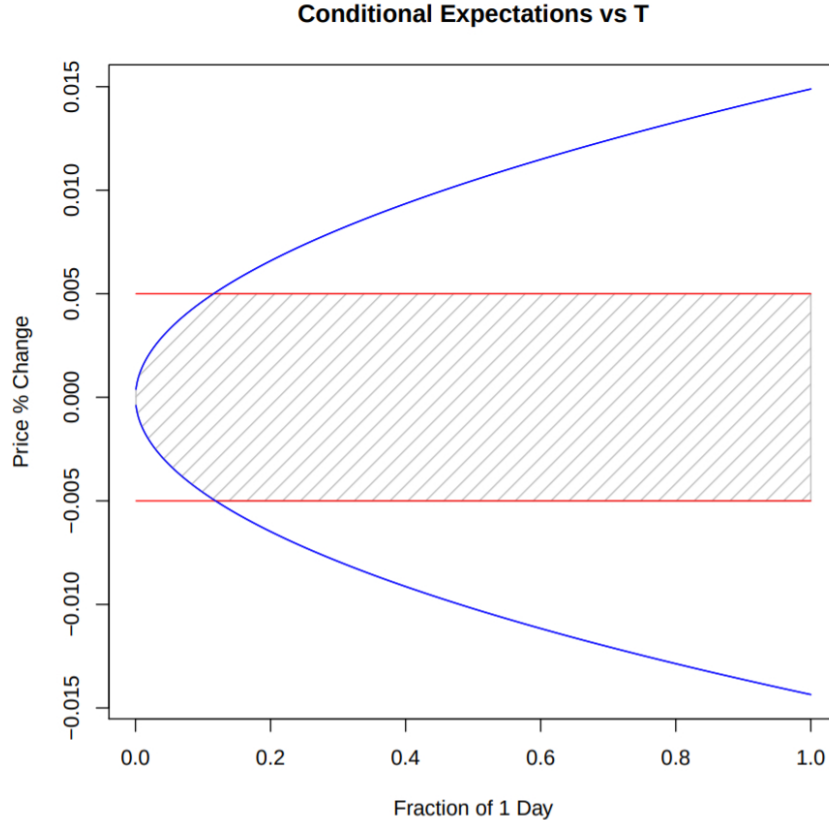
The dynamic around how to handle the LB and UB variables is in essence the entire core of this paper. Unsurprisingly we have built upon the innovative work from Chitra, Evans and Angeris [3] who eloquently provided a framework around optimal fees and arbitrage bounds for AMMs. They refer to this no arbitrage area as the "no-fee interval". In this model the arbitrage bounds are specific to each AMM and will be mainly dependent upon the fees the AMM charges for a swap as well as technically the gas price at the time of the swap. For the purposes of this paper though we will assume that the gas component of the LB and UB variables are de minimis and that the fee component dominates the implied arbitrage bound levels. One thing to call out is that LB and UB technically are unique to each arbitrage as DEX to DEX arbitrage will likely have higher cumulative fees involved in the transaction while a DEX to CEX arbitrage will only be paying one side of the swap fee on the DEX leg and a smaller fee on the CEX leg. Irregardless we assume that for a hypothetical AMM we only need to worry about an average transaction representing a mixture of all arbitrages which is likely mainly comprised of DEX to DEX trading. LB and UB are expressed in reference to 1 i.e. 1.005 would mean the upper bound of the no arbitrage zone is 0.5% higher than the current price and 0.995 would indicate that the lower bound is 0.5% lower than the current price.

This model attempts to represent the tradeoff between more frequent updating of the external oracle versus the expected cost of if an arbitrageur is able to capitalize upon a price difference between the AMM and the external market. The α parameter accounts for the relative risk that a DEX would expect of being arbitrated. For the majority of this section we effectively assume that arbitrage is omnipresent and that our hypothetical AMM needs to be as conservative as possible with accounting for it. A critical assumption underlying this model is that the amount of trading with the AMM is less frequent when compared with the broader market and thus prices in the internal LP become stale until either 1) an arbitrageur realigns the price with the external market or 2) we update the price via the external oracle.

To dive deeper into the formula 1 we have first the expected loss due to arbitrage followed by the cost of querying the oracle. For the former we weight the expected value by the likelihood of arbitrage occurring (α) and as mentioned earlier scale by a constant which is outside the scope of this paper to translate price changes into USD losses. The cost of querying the oracle is fairly self explanatory for as we increase our updating frequency, the associated cost in computational fees should increase too. With regards to the component representing expected loss due to arbitrage, the $\min\{0, UB - E[X|X \geq 0]\}$ and $\min\{0, E[X|X < 0] - LB\}$ pieces reflect how within the no fee interval arbitrage should never happen, assuming all arbitragers are rational utility maximizing participants. $E[X|X \geq 0]$ and $E[X|X < 0]$ are the conditional expected values for price increases and decreases respectively.

Armed with this formula we want to find the updating period t that maximizes the value (or minimizes the loss) we can get from including an external oracle. Unfortunately given the inclusion of the no arbitrage bound it is not as simple as setting the derivate of the formula with respect to t equal to 0 but we do identify that intuitively the problem reduces down to finding the smallest t that allows for $LB = E[X|X < 0]$ or $UB = E[X|X \geq 0]$. After finding the optimal t , the rest of the formula is used as a sanity check to make sure that it is even worth including an oracle updating service. For example, an especially small pool could find that the cost of maintaining this constant price updating could drain the liquidity providers of their funds. Or if gas costs were consistently very high it still may not make sense to depend upon an oracle at all.

Figure 1: Example of lognormal conditional expected values for increases and decreases in price (the blue lines) over one day with $\mu = 0.1$, $\sigma = 0.35$ and fees of 0.5%. The grey area is the no arbitrage zone.



Theorem 1 (Optimal Oracle Updating Frequency). *For a cryptocurrency's price changes that are assumed to follow $\text{Lognormal}(\mu, \sigma^2)$ and have no arbitrage bounds of LB and UB , the optimal updating frequency is:*

$$t = \min\{t_{UB}, t_{LB}\}$$

with

$$t_{UB} = \left(\frac{\frac{\beta\mu}{\sigma} - \beta UB(\frac{\mu}{\sigma} + \sigma) + \sqrt{(\frac{\beta\mu}{\sigma} - \beta UB(\frac{\mu}{\sigma} + \sigma))^2 + 16(\mu + \frac{\sigma^2}{2})(UB - 1)}}{4(\mu + \frac{\sigma^2}{2})} \right)^2$$

$$t_{LB} = \left(\frac{\beta LB(\frac{\mu}{\sigma} + \sigma) - \frac{\beta\mu}{\sigma} - \sqrt{(\beta LB(\frac{\mu}{\sigma} + \sigma) - \frac{\beta\mu}{\sigma})^2 + 16(\mu + \frac{\sigma^2}{2})(LB - 1)}}{4(\mu + \frac{\sigma^2}{2})} \right)^2$$

Proof. We start with finding the t that allows for $LB = E[X|X < 0]$. Given that X follows a lognormal distribution with mean μ and standard deviation σ we have:

$$\begin{aligned} LB &= E[X|X < 0] \\ &= e^{(\mu + \frac{\sigma^2}{2})t} \frac{\Phi(\frac{\ln(1) - \mu t - \sigma^2 t}{\sigma\sqrt{t}})}{\Phi(\frac{\ln(1) - \mu t}{\sigma\sqrt{t}})} \end{aligned} \quad (2)$$

Now using the logistic approximation for the normal CDF [Appendix A] we simplify

$$\begin{aligned} &= e^{(\mu + \frac{\sigma^2}{2})t} \frac{\frac{1}{1 + e^{-\beta(\frac{\mu t - \sigma^2 t}{\sigma\sqrt{t}})}}}{\frac{1}{1 + e^{-\beta(\frac{\mu t}{\sigma\sqrt{t}})}}} \\ &= e^{(\mu + \frac{\sigma^2}{2})t} \frac{1 + e^{-\beta(\frac{\mu t}{\sigma\sqrt{t}})}}{1 + e^{-\beta(\frac{\mu t - \sigma^2 t}{\sigma\sqrt{t}})}} \end{aligned} \quad (3)$$

Leveraging the first order taylor approximation of $e^x \approx 1 + x$ we get

$$= (1 + (\mu + \frac{\sigma^2}{2})t) \frac{2 + \beta(\frac{\mu t}{\sigma\sqrt{t}})}{2 + \beta(\frac{\mu t + \sigma^2 t}{\sigma\sqrt{t}})} \quad (4)$$

Expanding the terms results in

$$2(LB - 1) + (\beta LB(\frac{\mu}{\sigma} + \sigma) - \frac{\beta\mu}{\sigma})\sqrt{t} - 2(\mu + \frac{\sigma^2}{2})t - \frac{\beta\mu}{\sigma}(\mu + \frac{\sigma^2}{2})t^{1.5} = 0 \quad (5)$$

Now we can quickly recognize that introducing a change of variables of the form $t = x^2$ allows this to become a cubic polynomial that has a closed form solution. Unfortunately this closed form solution is extremely unwieldy and doesn't provide us with much intuition around how to interpret the relationship between the various variables (see [7]). Instead we look to leverage a technique from perturbation theory that will give us

a good local approximation while simplifying the solution dramatically [8]. We still proceed with the change of variables first and introduce simplifying variables to make the rest of the proof easier to follow:

$$\begin{aligned}
A_0 + A_1x - A_2x^2 - A_3x^3 &= 0 \\
A_0 &= 2(LB - 1) \\
A_1 &= (\beta LB(\frac{\mu}{\sigma} + \sigma) - \frac{\beta\mu}{\sigma}) \\
A_2 &= 2(\mu + \frac{\sigma^2}{2}) \\
A_3 &= \frac{\beta\mu}{\sigma}(\mu + \frac{\sigma^2}{2})
\end{aligned} \tag{6}$$

Within perturbation theory we are able to define our problem in terms of a power series that contains a small error term that measures the deviation from the true solution. In our case the approximate solution to our roots of x is $x_0 + \epsilon x_1 + \epsilon^2 x_2 + \dots$ with $\epsilon \ll 1$ (i.e. ϵ being significantly smaller than 1). We truncate this series at first order and have $x \approx x_0 + \epsilon x_1$ ($\epsilon \rightarrow 0$).

Now to simplify our original equation we recognize that $|A_3| \ll |A_i|$ with $i = \{0, 1, 2\}$ for most reasonable combinations of μ and σ (see Appendix B for more details). It is with the A_3 term that we now substitute in the ϵ term:

$$A_0 + A_1x - A_2x^2 - \epsilon x^3 = 0 \tag{7}$$

and now insert our approximation of x into the above equation

$$A_0 + A_1(x_0 + \epsilon x_1) - A_2(x_0 + \epsilon x_1)^2 - \epsilon(x_0 + \epsilon x_1)^3 = 0 \tag{8}$$

This equation is then expanded and we drop any terms with ϵ^n for $n > 1$ as the contribution of those terms are expected to be negligible.

$$A_0 + A_1x_0 - A_2x_0^2 + \epsilon(A_1x_1 - 2A_2x_0x_1 - x_0^3) = 0 \tag{9}$$

Now the perturbation series can be solved by setting each ϵ^n term for $n = \{0, 1\}$ equal to 0:

$$\begin{aligned}
\epsilon^0 : A_0 + A_1x_0 - A_2x_0^2 &= 0 \implies \\
: x_0 &= \frac{A_1 \pm \sqrt{A_1^2 + 4A_2A_0}}{2A_2} \\
\epsilon^1 : A_1x_1 - 2A_2x_0x_1 - x_0^3 &= 0 \implies \\
: x_1 &= \frac{x_0^3}{A_1 - 2A_2x_0}
\end{aligned} \tag{10}$$

Taking the solutions for x_0 and x_1 we substitute these two terms back into $x \approx x_0 + \epsilon x_1$ along with setting ϵ to A_3 again we get

$$x \approx \frac{A_1 \pm \sqrt{A_1^2 + 4A_2A_0}}{2A_2} + A_3 \frac{x_0^3}{A_1 - 2A_2x_0} \tag{11}$$

Recognizing that x_0 's root should be close to 0 we can further reduce to

$$x \approx \frac{A_1 - \sqrt{A_1^2 + 4A_2A_0}}{2A_2} + A_3 \frac{x_0^3}{A_1 - 2A_2x_0} \tag{12}$$

Since A_3 has been shown in Appendix B to be quite small under most circumstances and that x_0 's value is close to 0 (and by being taken to the 3rd power only magnifies this more) the second term can be dropped without losing much accuracy. Finally also accounting for the original change of variables that we did before ($t = x^2$) we arrive at

$$t \approx \left(\frac{A_1 - \sqrt{A_1^2 + 4A_2A_0}}{2A_2} \right)^2 \quad (13)$$

By swapping back our A_i variables from earlier we arrive at [1] listed above for LB 's optimal t . Following the same steps to find $UB = E[X|X \geq 0]$ is trivial. \square

One aspect to explicitly call out is that we are operating under the assumption that the AMM has an accurate estimation of the lognormal distribution's parameters. While this is by no means an easy task in practice, under normal situations the drift term tends to not carry too much weight in the above optimal t equations and almost all of the importance rests with the arbitrage bounds and σ . Below are a series of graphs illustrating the interaction between the 3 components. It is important to identify that the only scenarios where μ has any impact is when we are in a high μ and low σ regime which are contradictory for cryptocurrency markets. The crypto ecosystem is notoriously volatile and random large movements are the norm. This is in direct conflict with the combinations that would lead to μ being important for the calculation of t . Sample tables are provided in Appendix C with actual values of t .

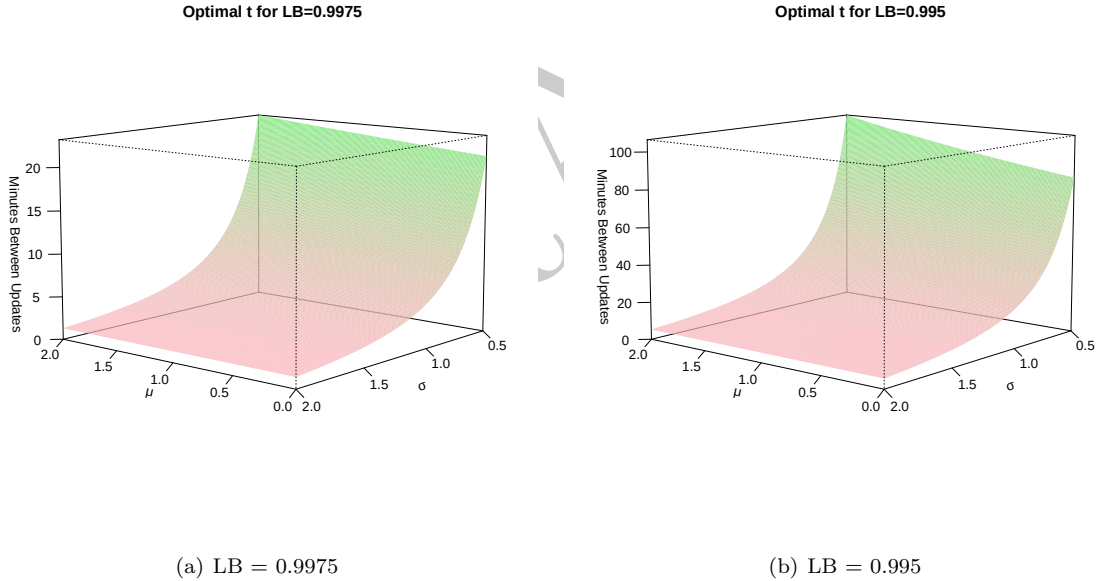


Figure 2: The implied optimal t in minutes across multiple combinations of μ , σ and LB

$$x = \sqrt[3]{\left(\frac{-b^3}{27a^3} + \frac{bc}{6a^2} - \frac{d}{2a}\right) + \sqrt{\left(\frac{-b^3}{27a^3} + \frac{bc}{6a^2} - \frac{d}{2a}\right)^2 + \left(\frac{c}{3a} - \frac{b^2}{9a^2}\right)^3}} + \sqrt[3]{\left(\frac{-b^3}{27a^3} + \frac{bc}{6a^2} - \frac{d}{2a}\right) - \sqrt{\left(\frac{-b^3}{27a^3} + \frac{bc}{6a^2} - \frac{d}{2a}\right)^2 + \left(\frac{c}{3a} - \frac{b^2}{9a^2}\right)^3}} - \frac{b}{3a}.$$

Figure 3: The unenjoyable solution to a cubic polynomial [7]

Implementation

While it is unrealistic to expect any sort of robust estimation of σ to be done on chain but there likely are simple regime classifying rules that can do a decent job of approximating the value of the volatility parameter. The no arbitrage bounds should stay fixed for any given pool so the estimation of σ is the only real area of concern. As Appendix C shows while there are certainly changes to t that occur as σ is varied, as long as we are in neighborhood of the correct σ value the estimated updating frequency should be reasonable. The AMM could be initialized with a set of optimal t s (calculated as the minimum of t_{UB} and t_{LB}) and dynamically update the selected t when necessary.

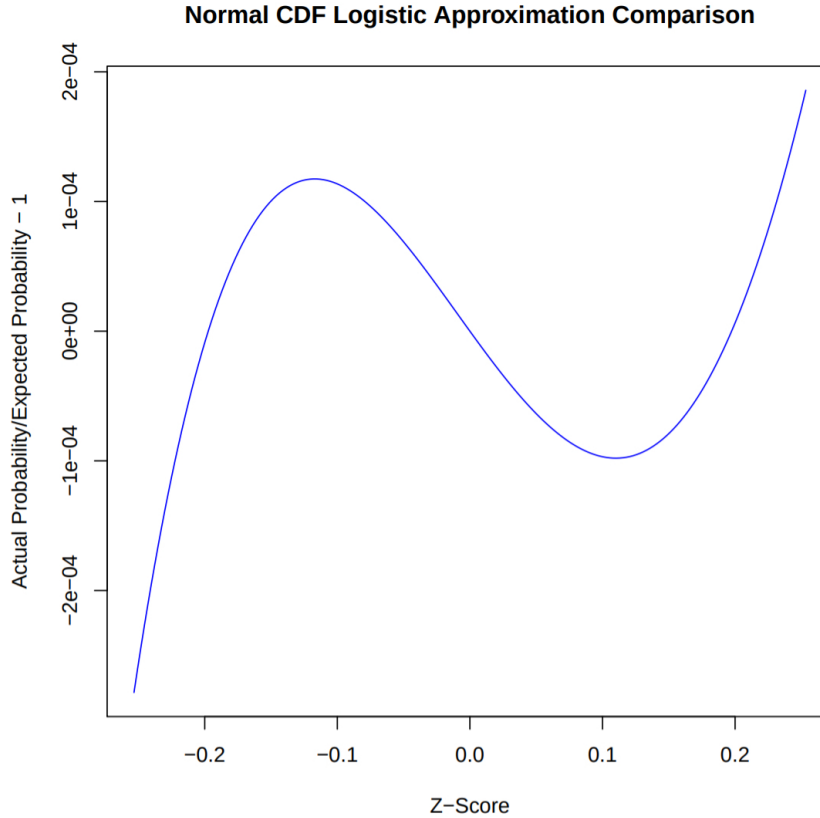
Conclusion

Within this paper we have found a simplified closed form solution for what the optimal updating time should be for a cryptocurrency under the lognormal distribution so as to minimize the impact of arbitrage. From this framework it can then be established around whether it is even economically worthwhile to include an oracle or not for an AMM. The optimal updating time appears to only be dependent on the size of the arbitrage free zone and the volatility of the cryptocurrency, the drift has little to no effect in most circumstances. We hope that this analysis proves to be a helpful starting point for any DEXes that are looking to optimize their usage of external oracles. Many areas of future research likely exist in this space around topics such as a much more complete economic model to quantify the pros/cons of frequent updating, leveraging different probability distributions to find the optimal updating period and more efficient approximations to some of the formulas derived earlier.

Appendices

A Normal CDF Approximation

The standard normal cumulative density function is one of the most well known distributions in all of statistics [4]. Despite the prevalence of this function and its widespread use, technically a closed form solution does not exist as it depends heavily on the error function $erf(x) = \frac{2}{\sqrt{\pi}} \int_0^x e^{-t^2} dt$ with the CDF being $\Phi(x) = \frac{1}{2}(1 + erf(\frac{x}{\sqrt{2}}))$. x is assumed to be normally distributed with a mean of 0 and a standard deviation of 1. To help with making the analysis more tractable we leveraged an approximation of the normal CDF through the usage of the logistic function $f(x) = \frac{1}{1+e^{-\beta x}}$ with β being the factor that controls the growth rate of the curve [5]. A calibration was ran to identify the ideal β that best matched a range of probabilities from 40% to 60%, which is the domain we would realistically expect this approximation to be used on. Within the analysis the probabilities in the range we iterated on were done uniformly so as to consistently sample in probability space, not z-score space which would have put extra emphasis on the tails. That being said it could be calibrated to the entire distribution or another specific section of it for a different problem. The optimal β of 1.5986 was found by minimizing the mean squared error between the true normal CDF and the logistic approximation. Additional checks were done over this range around using mean absolute percentage error as our minimizing metric but it yielded an almost identical β . Below is a graph showing the percentage difference between the actual normal CDF and the approximated normal CDF.



B A_3 Magnitude

For the perturbation expansion and ϵ replacement to hold we need to first establish that $|A_3| \ll |A_i|$ with $i = \{0, 1, 2\}$

$$\begin{aligned} A_0 &= 2(LB - 1) \\ A_1 &= \beta LB \left(\frac{\mu}{\sigma} + \sigma \right) - \frac{\beta\mu}{\sigma} \\ A_2 &= 2\left(\mu + \frac{\sigma^2}{2}\right) \\ A_3 &= \frac{\beta\mu}{\sigma} \left(\mu + \frac{\sigma^2}{2} \right) \end{aligned} \tag{14}$$

Starting first with A_2 it is apparent that A_3 is just a scaled version of A_2 with the additional multiplier's magnitude of $\frac{\beta\mu}{\sigma}$ needing to be less than 2 for $A_3 \ll A_2$ to hold. Given that β is around 1.5986 [see Appendix A] and σ is almost surely greater than μ we have proven this first comparison.

For $A_3 \ll A_1$ we will proceed as follows:

$$\begin{aligned} A_3 &\ll A_1 \\ \frac{\beta\mu}{\sigma} \left(\mu + \frac{\sigma^2}{2} \right) &\ll \beta LB \left(\frac{\mu}{\sigma} + \sigma \right) - \frac{\beta\mu}{\sigma} \\ \mu + \frac{\sigma^2}{2} &\ll \frac{LB\sigma}{\mu} \left(\frac{\mu}{\sigma} + \sigma \right) - 1 \\ \mu + \frac{\sigma^2}{2} &\ll LB \left(1 + \frac{\sigma^2}{\mu} \right) - 1 \\ \mu + \frac{\sigma^2}{2} + 1 &\ll LB + \frac{LB\sigma^2}{\mu} \\ 0 &\ll -\mu^2 + \mu \left(LB - \frac{\sigma^2}{2} - 1 \right) + LB\sigma^2 \end{aligned} \tag{15}$$

In almost all instances $LB \approx 1$ so we can further simplify and solve the quadratic formula for μ in terms of σ

$$\mu \ll -\frac{\sigma^2}{4} + \frac{\sigma}{2} \sqrt{\frac{\sigma^2}{4} + 4} \tag{16}$$

So as long as this relationship holds (and in almost all realistic instances it does) then $A_3 \ll A_1$ holds too.

We also note that while A_0 is also quite small, swapping it out for the perturbation parameter ϵ instead of A_3 leads to implied t values that are non sensical (negative t) so our approach to allow the cubic term's coefficient be substituted for ϵ is the least bad option.

C μ , σ and LB Combinations

Lower Bounds	μ	σ	Optimal t (minutes)
0.9975	0.25	1.00	5.205
0.9975	0.50	1.00	5.222
0.9975	0.75	1.00	5.239
0.9975	1.00	1.00	5.256
0.9975	0.25	1.25	3.327
0.9975	0.50	1.25	3.334
0.9975	0.75	1.25	3.341
0.9975	1.00	1.25	3.348
0.9975	0.25	1.50	2.309
0.9975	0.50	1.50	2.312
0.9975	0.75	1.50	2.316
0.9975	1.00	1.50	2.319
0.9975	0.25	1.75	1.696
0.9975	0.50	1.75	1.698
0.9975	0.75	1.75	1.699
0.9975	1.00	1.75	1.701
0.9975	0.25	2.00	1.298
0.9975	0.50	2.00	1.299
0.9975	0.75	2.00	1.300
0.9975	1.00	2.00	1.301
0.9950	0.25	1.00	21.078
0.9950	0.50	1.00	21.218
0.9950	0.75	1.00	21.359
0.9950	1.00	1.00	21.503
0.9950	0.25	1.25	13.458
0.9950	0.50	1.25	13.515
0.9950	0.75	1.25	13.572
0.9950	1.00	1.25	13.630
0.9950	0.25	1.50	9.334
0.9950	0.50	1.50	9.361
0.9950	0.75	1.50	9.389
0.9950	1.00	1.50	9.416
0.9950	0.25	1.75	6.852
0.9950	0.50	1.75	6.867
0.9950	0.75	1.75	6.882
0.9950	1.00	1.75	6.897
0.9950	0.25	2.00	5.244
0.9950	0.50	2.00	5.252
0.9950	0.75	2.00	5.261
0.9950	1.00	2.00	5.270

References

- [1] DODO Team. "DODO Whitepaper" *GitHub* <https://dodoex.github.io/docs/docs/whitepaper/>
- [2] Finematics. "What is Bancor V2? Chainlink Integration and Dynamic Automated Market Maker Explained." *Medium (blog)*, August 10, 2020. <https://finematics.medium.com/what-is-bancor-v2-chainlink-integration-and-dynamic-automated-market-maker-explained-2104049f7e73>.
- [3] Alex Evans Guillermo Angeris Tarun Chitra, 2021. "Optimal Fees for Geometric Mean Market Makers," Papers 2104.00446, arXiv.org.
- [4] Wikipedia contributors, "Normal distribution," Wikipedia, The Free Encyclopedia, https://en.wikipedia.org/w/none?title=Normal_distribution&oldid=1047324524 (accessed October 2, 2021).
- [5] Wikipedia contributors, "Logistic function," Wikipedia, The Free Encyclopedia, https://en.wikipedia.org/w/none?title=Logistic_function&oldid=1045492244 (accessed October 2, 2021).
- [6] Wikipedia contributors, "Log-normal distribution," Wikipedia, The Free Encyclopedia, https://en.wikipedia.org/w/none?title=Log-normal_distribution&oldid=1046866551 (accessed October 1, 2021)
- [7] Vanderbilt. (n.d.). The Cubic Formula. The cubic formula. Retrieved October 2, 2021, from <https://math.vanderbilt.edu/schectex/courses/cubic/>.
- [8] Bender, Carl M., and Steven A. Orszag. 1978. Advanced mathematical methods for scientists and engineers. New York: McGraw-Hill.