*22[th] International Conference on Modeling & Applied Simulation*
20[th] International Multidisciplinary Modeling & Simulation Multiconference

2724-0037 © 2023 The Authors.
doi: 10.46354/i3m.2023.mas.012

# Pythagorean triples for linear approximations of probabilistic production streams

Ivo Formánek[1], David Kruml[2,*], Jan Paseka[2] and Jan Šťovíček[2]

[1]Department of Entrepreneurship and Management, Pan-European University, Spálená 76/14, Prague, 110 00, Czech Republic
[2]Department of Mathematics and Statistics, Masaryk University, Kotlářská 2, Brno, 611 37, Czech Republic

[*]Corresponding author. Email address: kruml@math.muni.cz

## Abstract

Today's global hypercompetitive environment considerably demands optimizing the activities of all companies that want to be successful in the long term. Several tools are available on the market for optimizing corporate activities. While these tools are powerful, they are costly and challenging to operate. However, the tools are not suitable for small and medium-sized enterprises (SMEs). SMEs need tools powerful enough but at the same time user-friendly and cost-effective. The authors of this article are involved in developing one of these tools suitable for SMEs. This paper presents a part of the developed tool that applies a method based on the properties of Pythagorean triples for the numerical modeling of probabilistic flows. The resulting linear approximations are expressed in terms of rational coefficients, which makes them practical for possible implementation.

**Keywords**: Value stream; probabilistic production stream; linear approximation; Pythagorean triple

## 1. Introduction

There is enormous pressure on the optimal management of companies' critical processes. Optimal setup and management of the key processes determine the successful operation of companies in the market. The most common optimality criteria are almost always minimum operating costs and maximum business profit. When we optimize critical processes, their individual parts must also be defined, identified, and optimized appropriately. Only from optimally configured sub-processes can optimally working process chains characterized by high efficiency and effectiveness be assembled. It is almost impossible to encounter a situation where process chains do not change their topology or parameters in the long term. Reacting flexibly to changes caused by changing technical, organizational, or business conditions is almost always necessary.

In addition, the chains also include various warehouses, interim storage facilities, etc. (from here on stacks only), which also change their status. In addition, the smaller the stack capacity, the more difficult it is to manage process chains optimally. An extreme situation arises when process chains operate entirely without stacks, e.g., in Just in Time (JIT) mode.

The situation of prominent and wealthy companies is relatively straightforward when optimizing company processes. Large companies can afford to invest heavily in building and equipping specialized teams to implement optimization. In the case of SMEs, however, the situation is different. SMEs usually do not have sufficient resources to employ optimization specialists or to purchase powerful but also costly and difficult-to-operate optimization tools. SMEs need powerful yet inexpensive and user-friendly tools that non-specialists can operate. The authors of the submitted article are involved in developing one of these tools suitable for SMEs.

The tool preferably applies the Value Chain (VC) concept defined in Porter (1985) for the optimal management of process chains. The VC model allows a better understanding of the cost behavior in chains and, consequently, better identification of potential Value Added (VA) sources. In particular, the VC model notes strategically essential activities such as research, design, production, sales, delivery, and after-sales service described in Porter (1985); Kotler et al. (2021), which are critical to the success of the firm. We also preferably apply Value Stream Mapping (VSM) (see Liker (2020); Rother and Shook (2003)) to identify VCs with VA resources.

Managing VCs in the case of simple process links is relatively straightforward. However, a considerably more complex situation arises when many interacting activities need to be addressed whose interrelations and interactions are complicated.

The situation is further complicated when management has to react to events that could not have been foreseen (e.g., unexpected technical failures, supply interruptions, etc.). In this case, finding the most economically acceptable solution using a simulation model is advantageous. Several planning systems in the market can be used for optimal chain management. However, few can dynamically react to the situation and simultaneously use Big Data from VC, which is required today.

Streams in mass production could have specific dispersion. In our paper, we derive a method based on the properties of Pythagorean triples for the numerical modeling of such a probabilistic stream.

The article is organized as follows: In Section 2, we recall the distinction between the perfect production stream and its linear idealization, and the probabilistic production stream with selected quantile and its idealization employing a parabolic curve. We also mention here some basic properties of probabilistic production streams.

Section 3 explains the background to the solution methodology, followed by the details regarding different constructions for creating the Pythagorean triples. In particular, inner and outer linear approximation methods based on Pythagorean triples are described. Section 4 starts with an illustrative example of 100 runs of a simple production stream and we discuss possible modifications to our approach here. Finally, Section 5 concludes the paper.

## 2. State of the art

The study of probabilistic data streams raises a host of new, exciting research challenges for approximating discrete production streams. One of them is the use of linear approximations. Linearization is a significant step in our approach. It can be considered a nonlinearity correction between the output and the associated measured quantity. Usually, the nonlinearity can be reduced by using linearization schemes or a linearization algorithm, as in Chen and Wang (2013); Luo et al. (2015).

The construction of linear approximations of quadratic functions has been an active area of research over the past few decades (see, e.g., Pottmann et al. (2000)) with emphasis on the optimality of the approximation. Since we are in our contribution motivated by practical requirements for immediate response, we proceed differently.

Every discrete production stream can be described as a broken line in a mass-time space. A move along the mass axis represents an increase in quantity, while a movement along the time axis represents a delay between two subsequent events. Suppose an iterative "perfect" process creates the stream. In that case, all the steps are equal, and the graph of the stream fits to a parallelogram (see Fig. 1). The tangents can be interpreted as perfect continuous streams approximating the given discrete stream from above and below.

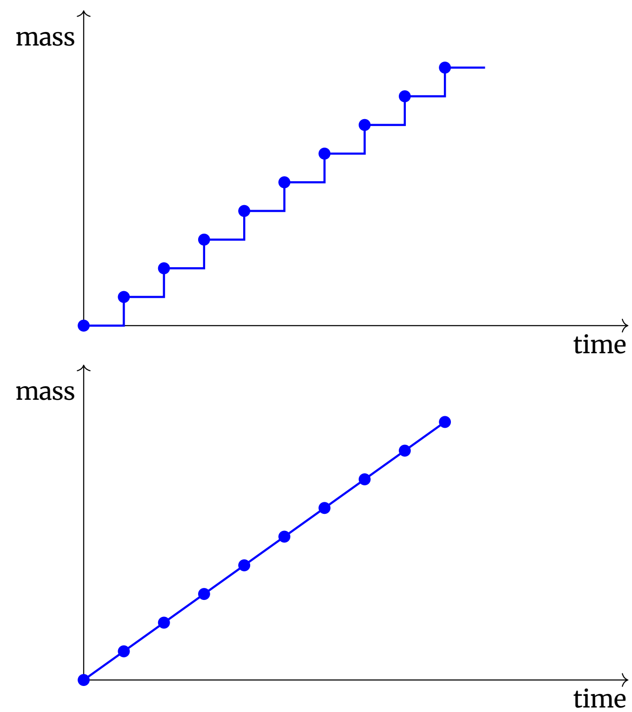


**Figure 1.** Perfect production stream (above) and linear idealization of its upper bound (below).

In (Kruml and Paseka, 2018), we have shown that linear approximations are very convenient for modeling a stack development. We add the approximating linear functions if a stack is fed and drawn by more processes. The algorithm seeks eventual collisions on the stack. It works recursively and improves the approximations only for intervals where it did not get identical results for the upper and lower bounds. The algorithm is effective, especially for highly regular production, when processes are performed in many repetitions. (The resulting graph has a smaller number of intervals of activity/non-activity and a relatively narrow approximating parallelogram.)

In reality, processes are often far from such perfectness, and one has to consider some uncertainty in production parameters. The time parameters, e. g., a cycle time, could have a dispersion, or the process could fail in quality, and the product must be discarded. Such phenomena provide changes in the perfect shape of the graph — the steps need not be of equal width, and some of them could be missing. Since possible stream runs are no longer unique, we have to consider a probabilistic environment.

While in the perfect case, we ask whether all input and output streams of a stack are or are not in a collision, in the probabilistic environment, we ask if a collision may or may not happen with a certain probability. That is, we test the plan merely concerning given reliability. For such a purpose, we need to model quantiles of the probabilistic environment (see Fig. 2).

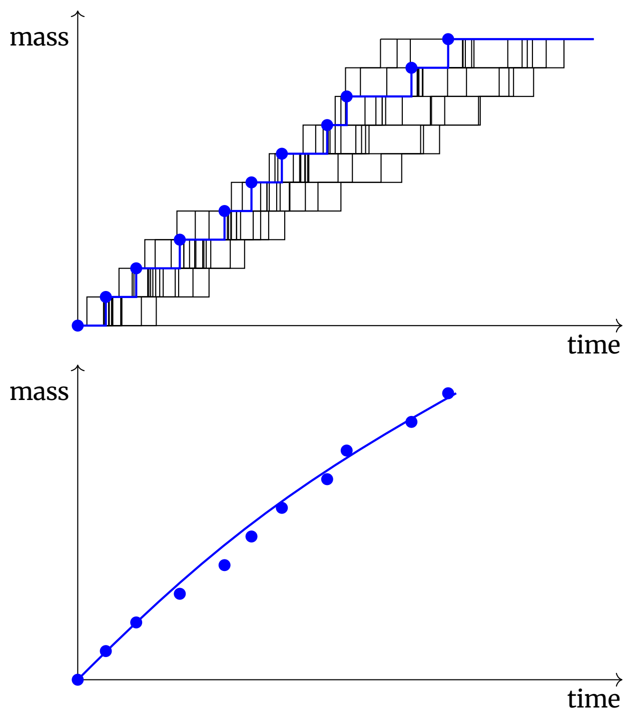


**Figure 2.** A probabilistic production stream with selected quantile (above) and its idealization by means of a parabolic curve (below).

In (Emir et al., 2019) and (Emir et al., 2021), we studied such probabilistic streams and suggested simplified calculus for their addition. Likewise, in those works, we still consider the random effects independent, i.e., that cycles of the process have "no memory" and are not affected by previous cycles. Of course, some unexpected events could significantly impact production (worker's experience, wrong setting of a machine). Still, such a case can be modeled as uncertainty on a higher plan level — the process has more instances, and some of them are randomly chosen for specific periods.

Finally, we have also assumed that the random variables have some natural probability distribution and can be represented merely by three parameters $\mu, \sigma, \gamma$ where $\mu$ is the mean, $\sigma$ is the standard deviation, and $\gamma$ a parameter concerning skewness. We have verified on several examples that those three parameters and calculus could be sufficient for modeling probability in practice. In the later work (Emir et al., 2021), we have replaced $\gamma$ by parameter $\gamma_p$ to describe the correction of the $p$th quantile concerning a so-called reference distribution which was some of "natural" symmetric distributions (normal, logistic). It turned out that triples $\mu, \sigma, \gamma_p$ can be processed in a similar fashion like $\mu, \sigma, \gamma$ and we can quite easily predict the development of the $p$th quantile in mass-time. In particular, the parameters $\gamma$ and $\gamma_p$ are invariant to the summing of equal independent random variables.

Under such assumption, the stream fulfils the following facts or conclusions:

- Moves between events of the production are given by equal independent random variables $X_1, X_2, \ldots X_n$
- The global effect of the stream is given as a sum of random variables $Y = X_1 + X_2 + \cdots + X_n$.
- Sum of means is a mean of the sum, i. e.,

$$\mu_Y = \mu_{X_1} + \mu_{X_2} + \cdots + \mu_{X_n} = n\mu_{X_1}. \tag{1}$$

- Sum of variances is a variance of the sum, i. e.,

$$\sigma_Y^2 = \sigma_{X_1}^2 + \sigma_{X_2}^2 + \cdots + \sigma_{X_n}^2 = n\sigma_{X_1}^2. \tag{2}$$

- For a standard deviation this yields:

$$\sigma_Y = \sqrt{n\sigma_{X_1}^2} = \sqrt{n}\sigma_{X_1}. \tag{3}$$

- In normal, logistic, and some other probabilistic distributions, the $p$th quantile can be expressed as

$$\mu + z_p\sigma$$

  for suitable $z_p$.
- If $p$th quantile of random variable $X_1$ is

$$\mu_{X_1} + z_p\sigma_{X_1} + \gamma_p$$

  for suitable $z_p, \gamma_p$, then $p$th quantile of random variable $Y$ can be expected as

$$n\mu_Y + z_p\sigma_Y + \gamma_p = n\mu_{X_1} + \sqrt{n}z_p\sigma_{X_1} + \gamma_p. \tag{4}$$

While equations (1), (2), and (3) are well-known properties of random variables, quantile guess (4) is approximate and based on an observation that $\gamma_p$ remains constant for any $n$. This fact is explained and verified in (Emir et al., 2021).

For simplicity, let us assume that the randomness concerns only cycle time. We can conclude that the quantiles should develop along parabolic curves of the form

$$t = n\mu + \sqrt{n}z_p\sigma + \gamma_p. \tag{5}$$

If we need to sum parallelly performing streams, we must solve the equation, i. e. find the quantity $n$ produced in time $t$ (with respect to the $p$th quantile). This yields

$$(t - \gamma_p - n\mu)^2 = nz_p^2\sigma^2 \tag{6}$$

and after substitution

$$a = \left(\frac{z_p\sigma}{\mu}\right)^2, \qquad b = \frac{t-\gamma_p}{\mu}$$

we get equation

$$n^2 + (-a - 2b)n + b^2 = 0 \tag{7}$$

with roots

$$n_{1,2} = \frac{a + 2b \pm \sqrt{a(a+4b)}}{2}. \tag{8}$$

Hence the square root operation appears in expressions (5) and (8). The summation of such functions is non-trivial and could increase the computational complexity of calculations. We aim to find suitable linear approximations of the parabolic curves, which would provide summation methods similar to the model of perfect production.

## 3. Materials and Methods

We can see that in expression (8) there are both arithmetic and geometric means of $a$ and $a + 4b$:

$$A(a, a+4b) = \frac{a + a + 4b}{2} = a + 2b, \tag{9}$$
$$G(a, a+4b) = \sqrt{a(a+4b)}. \tag{10}$$

Our aim is to replace the square root expression $\sqrt{a(a+4b)}$ by a linear approximation in variables $a, b$. After that we would get

$$n_{1,2} = \alpha a + \beta b \tag{11}$$

for suitable coefficients $\alpha, \beta$. Since only $b$ is a linear function of time $t$, this would yield a linear correspondence between $n$ and $t$.

There are many possibilities for approximating $\sqrt{a(a+4b)}$, but among those, we will prefer partitions where the coefficients $\alpha$ and $\beta$ are rational. In such a case, further numerical calculations can be performed precisely with no risk of rounding errors.

By Euclid theorem, the expression $\sqrt{a(a+4b)}$ is an altitude of the right triangle, which splits the hypotenuse to $a$ and $a + 4b$ (see Fig. 3). Changing the proportion between $a$ and $a + 4b$, the right angle corner moves on Thales circle with the ratio $a + 2b$. As $t \to \infty$, parameter $b$ increases, and the smaller angle narrows.

The rational ratio of $\sqrt{a(a+4b)}$ to $a, b$ is obtained pre-

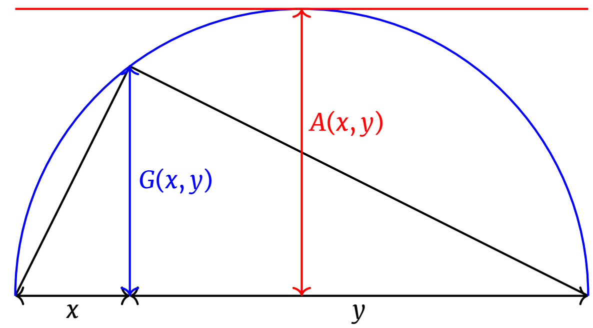


**Figure 3.** Visualization of arithmetic and geometric means in Thales circle.

cisely when the triangle is similar to a triangle with all edges integer. Such triangles are called *Pythagorean*, and their edges form *Pythagorean triples*. (Conversely, a triple of integers $A, B$ and $C$ is Pythagorean if $A^2 + B^2 = C^2$, that is, it makes a right triangle.) Thus we are looking for a sufficiently large and dense set of Pythagorean triples. In the unbounded case, the parabola is approximated by infinitely many linear sections. Hence we need methods for constructing infinite sequences of Pythagorean triples with decreasing angles.

### 3.1. Families of Pythagorean triples

The first such family of Pythagorean triples was allegedly discovered by Pythagoras himself as a sequence of triples of the form

$$A = 2k + 1, \qquad B = 2k^2 + 2k, \qquad C = 2k^2 + 2k + 1 \tag{12}$$

for integer $k$. The formula generates triples

$$\begin{array}{lll} (3,4,5), & (5,12,13), & (7,24,25), \\ (9,40,41), & (11,60,61), & (13,84,85), \\ (15,112,113), & (17,144,145), & (19,180,191), \\ (21,220,221), & (23,264,265), & \ldots \end{array}$$

Another family given by rules

$$A = 2k, \qquad B = k^2 - 1, \qquad C = k^2 + 1 \tag{13}$$

for integer $k \geq 2$ was found by Plato. This yields triples

$$\begin{array}{lll} (4,3,5), & (6,8,10), & (8,15,17), \\ (10,24,26), & (12,35,37), & (14,48,50), \\ (16,63,65), & (18,80,82), & (20,99,101), \\ (22,120,122), & (24,143,145), & \ldots \end{array}$$

Both families satisfy the expectation that the ratio $A : C$ (or the angle) decreases as $k \to \infty$. Notice that the Plato family doubles the density of the Pythagoras family in the sense that every second triple of Plato is also that of Pythagoras.

Later on, all Pythagorean triples were found by Euclid. Today, there are many constructions for creating the Pythagorean triples. We will use the enumeration principle discovered in McCullough and Wade (2003):

**Theorem 1.** *For each $h = pq^2$ with $p$ square-free let $d = 2pq$ for $p$ odd and $d = pq$ for $p$ even. Then for every pair $(h, k)$ of integers the triple*

$$A = h + dk, B = dk + \frac{(dk)^2}{2h}, C = h + dk + \frac{(dk)^2}{2h} \quad (14)$$

*is Pythagorean, and it is generated uniquely.*

We obtain the Pythagoras and Plato families of triples when $h = 1$ $(d = 2)$ or $h = 2$ $(d = 2)$, respectively, and $k$ runs over all integers. For $h = 4$ $(d = 4)$, we would get the Plato family with doubled coefficients. Still, $h = 8$ $(d = 4)$ yields a new family that extends the Plato family in the same fashion as the Plato family extends the Pythagoras family.

More generally, we observe that the density of triples doubles in each increment of $h$ in sequence $1, 2, 8, 32, \dots$. Hence, we will generate families with

$$h_i = \begin{cases} 1, & i = 1, \\ 2^{2i-3}, & i > 1 \end{cases} \quad (15)$$

and

$$d_i = \begin{cases} 2, & i = 1, \\ 2^{i-1}, & i > 1. \end{cases} \quad (16)$$

In what follows, let $T_i(k)$ be the triple given by $(h_i, k)$ and $T_i$ be the family of all $T_i(k)$ with $k$ integer.

### 3.2. Linear approximations

Now we will express the altitude $\sqrt{a(a + 4b)}$ of a triangle similar to a Pythagorean triple. We get equation

$$\sqrt{a(a + 4b)} = F(a + 2b) \quad (17)$$

where $F$ is a ratio between leg $A$ and hypotenuse $C$. In McCullough–Wade enumeration, it is given by

$$F = \frac{h + dk}{h + dk + \frac{(dk)^2}{2h}}. \quad (18)$$

Equation (17) results in a quadratic equation with only one positive (and reasonable) root

$$b = \frac{a}{2F^2}\left(1 - F^2 + \sqrt{F^2 + 1}\right), \quad (19)$$

and after substituting for $F$, we infer

$$b = \frac{dk(dk + 2h)}{4h^2}a. \quad (20)$$

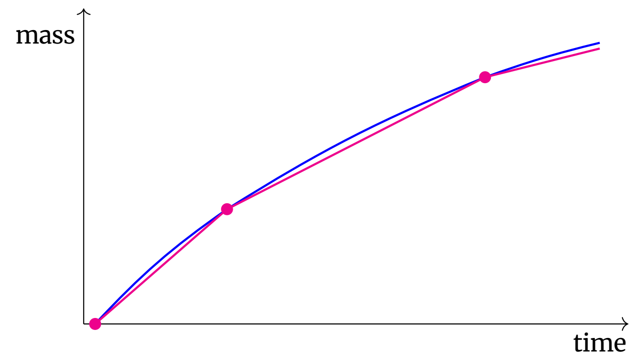


**Figure 4.** The inner linear approximation (magenta) simply connects points provided by a certain Pythagorean family.

The value of the point in function $f(b) = \sqrt{a(a + 4b)}$ is

$$f(b) = \frac{dk + h}{h}a. \quad (21)$$

Thus we get the point at which coordinates are rational concerning parameters $a$ and $b$, as desired.

In an *inner approximation* of the parabolic graph, we connect such consecutive points for some family of Pythagorean triples with fixed $h, d$ (see Fig. 4). The lines have the left or right endpoints given by $k - 1$ or $k$, respectively. After substitution and further improving we infer

$$\alpha = \frac{2hdk - hd + 2h^2 + (dk)^2 - d^2k}{h(d(2k - 1) + 2h)}, \quad (22)$$

$$\beta = \frac{4h}{2h + d(2k - 1)} \quad (23)$$

for the line

$$y_k = \alpha a + \beta b. \quad (24)$$

A maximal error of the approximation is calculated as a maximum of function

$$g(b) = \sqrt{a(a + 4b)} - (\alpha a + \beta b). \quad (25)$$

Since $g$ is continuous, non-negative, and differentiable in all points except the zero-points, the local maxima are realized in that $b_{max}$ that $g'(b_{max}) = 0$. (Notice that every interval of the partition has precisely one local maximum.) We infer that

$$b_{max} = \frac{ad(2k - 1)(d(2k - 1) + 4h)}{16h^2} \quad (26)$$

and the maximum error is

$$g(b_{max}) = 11ad(2k - 1) - \frac{a(8d^2k^2 - 8d^2k + d^2 + 16dhk - 8dh + 8h^2)}{4h(d(2k-1) + 2h)}. \quad (27)$$

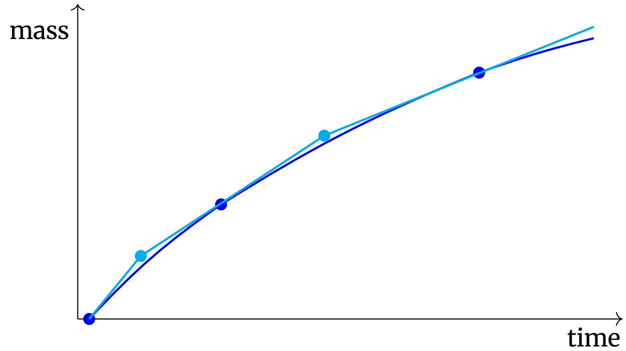


**Figure 5.** In the outer linear approximation (cyan), we find tangents at "Pythagorean points" (blue) and their intersections.

The global maximum always appears in the first interval ($k = 1$) and is equal to

$$g(b_{max}) = \frac{ad^2}{4h(d + 2h)}. \tag{28}$$

In an *outer approximation*, we use the points referring to Pythagorean triples as tangent points of the lines (see Fig. 5). The tangent direction is given by derivation of $\sqrt{a(a + 4b)}$, and the endpoints of lines are calculated as intersections of pairs of consecutive tangents. Assume that equations (20) and (21) are satisfied. Then

$$f'(b) = \frac{2h}{h + dk}, \tag{29}$$

and consequently, we infer pair of coefficients

$$\alpha = \frac{2hdk + 2h^2 + (dk)^2}{2h(h + dk)}, \tag{30}$$

$$\beta = \frac{2h}{h + dk} \tag{31}$$

for the tangent (24).

By substitution $k - 1$ for $k$, we find the previous tangent with coefficients, say $\bar{\alpha}$, $\bar{\beta}$, and solve equation

$$\alpha a + \beta b = \bar{\alpha} a + \bar{\beta} b. \tag{32}$$

The equation (32) yields a left endpoint of the approximating line in

$$b = \frac{ad(dk^2 - dk + 2hk - h)}{4h^2} \tag{33}$$

for $k \geq 2$. The case $k = 1$ is exceptional because there is no preceding tangent. Here we put $b = 0$. Substitution $k + 1$ instead of $k - 1$ would provide the right endpoint.

Contrary to the inner approximation, we obtain the local maxima of the difference function $g$ in the endpoints of lines. This result yields

$$g(b_{max}) = \frac{a}{h}\left(dk + h - \frac{1}{2} - \sqrt{d^2(k-1)k + dh(2k-1) + h^2}\right) \tag{34}$$

and again, the global maximum, i.e., the most significant error of the approximation, is reached at the left endpoint of the first line

$$g(0) = a\frac{d^2}{2h(h + d)}. \tag{35}$$

## 4. Results and Discussion

It remains to apply the general form of approximations to the case of Pythagorean families $T_i$. Substitution of $d_i, h_i$ from equations (15) and (16) provides further simplification, and we infer in Table 1 and Table 2.

**Table 1.** Inner approximation.

| inner approximation | |
|---|---|
| left endpoint | $b = a\frac{(k-1)(k-1+2^{i-1})}{2^{2i-2}}$, |
| right endpoint | $b = a\frac{k(k+2^{i-1})}{2^{2i-2}}$, |
| coefficient $\alpha$ | $1 + \frac{k(k-1)}{2^{i-2}(2k-1+2^{i-1})}$ |
| coefficient $\beta$ | $\frac{2^i}{2k-1+2^{i-1}}$ |
| maximal error | $\frac{a}{2^i(1+2^{i-1})}$ |

**Table 2.** Outer approximation.

| outer approximation | |
|---|---|
| left endpoint | $b = \begin{cases} 0, & k = 1, \\ a\frac{k^2+(2^{i-1}-1)k-2^{i-2}}{2^{2i-2}}, & k > 1, \end{cases}$ |
| right endpoint | $b = a\frac{k^2+(2^{i-1}+1)k-2^{i-2}}{2^{2i-2}}$, |
| coefficient $\alpha$ | $1 + \frac{k^2}{2^{i-1}(k+2^{i-2})}$ |
| coefficient $\beta$ | $\frac{2^{i-1}}{k+2^{i-2}}$ |
| maximal error | $\frac{a}{2^{i-1}(1+2^{i-2})}$ |

### 4.1. Example

Let us consider a simple production stream that produces a single item in a cycle, and the cycle time takes normal distribution of probability $N(\mu, \sigma^2)$ with $\mu = 1, \sigma = 0.3$. We assume that the production was performed 100 times, and each such batch consists of 100 repetitions of the process. We wish to model a 5%-quantile, i.e., a fictional stream for which 5% of runs are faster and 95% are slower.

The quantile can be expressed as $\mu - z_p\sigma$ with $z_p = 1.645$, hence the $n$th item should be produced in time

$$t = n\mu - \sqrt{n} z_p \sigma. \qquad (36)$$

This yields a continuos function in $n$ corresponding to an upper bound of "ideal random stream", that is the limit case that we would get for infinitely many batches. (The lower bound is obtained by subtracting 1 item from $n$.) We get

$$a = \left(\frac{z_p \sigma}{\mu}\right)^2 \approx 0.2435 \qquad b = \frac{t}{\mu} = t \qquad (37)$$

In reality, the 100 runs of the stream would be the only data. The quantile would be constructed by rearranging the field $100 \times 100$ to lists for each $n$ and choosing the 5th values for each such list. The statistical error has a standard deviation given by coefficient $s = 1/\sqrt{100} = 1/10$. Since $a$ depends linearly on variance $\sigma^2$, its statistical standard deviation is $as^2$. We can accept an approximation as "reliable" if its maximal error does not exceed the statistical error, that is,

$$g(b_{max}) \leq as^2. \qquad (38)$$

For Pythagorean families $T_i$, and the inner or outer approximation we infer

$$\frac{1}{2^i(1+2^{i-1})} \leq \frac{1}{100}, \qquad \frac{1}{2^{i-1}(1+2^{i-2})} \leq \frac{1}{100}, \qquad (39)$$

respectively. In the former case we get a minimal solution $i = 4$, in the later case $i = 5$. We expect finishing the batch in time

$$t = 100 - 10 \cdot 1.645 \cdot 0.3 = 95.07. \qquad (40)$$

The largest endpoints of approximations satisfy

$$95.07 \leq 0.2435 \cdot \frac{k(k+8)}{64}, \qquad (41)$$

$$95.07 \leq 0.2435 \cdot \frac{k^2 + 17k - 8}{256}. \qquad (42)$$

The minimal solutions are $k = 75$ or $k = 148$, respectively. Hence, the approximations correspond to families $T_4(k), 1 \leq k \leq 75$ or $T_5(k), 1 \leq k \leq 148$, respectively.

### 4.2. Discussion

The endpoints of approximations can be either precised by adding further points or relaxed by their omitting. The Pythagorean families $T_i$ provide a scheme where point density doubles by incrementing $i$ and the old points are still used. Since the approximation error is most significant at the beginning of stream, one could also consider "hybrid" families of Pythagorean triples where density of points decrease in time.

We have started with Pythagorean family $T_1$ but we can also go backwards and define more relaxed families $T_i$ for $i < 1$ by omitting all odd points in every step. Interestingly, it seems that all rules and expressions derived for $i \geq 1$ are still preserved.

The example shows that selected density of points seems to be redundant with respect to the size (and quality) of the dataset. (In the outer approximation we got more points than it came from the dataset.) Perhaps, one could consider a more relaxed ratio between the statistical error and the error of approximation. Also, we have measured the approximation error as a maximum of difference function $g$, but expressing it in another norm of $g$ might be more relevant.

## 5. Conclusion

In general, the choice of model has a significant impact on its quality and performance: the increased complexity of the model gives us higher fidelity to the underlying data and therefore higher accuracy but requires additional computational and I/O costs. These high cost naturally raises the question of whether highly sophisticated models are worth it.

The described method provides a way how to represent a certain type of random stream by polygonal approximations parametrized by rational coefficients. It can be used in the case that the stream is modeled stochastically (by a certain probability distribution) or statistically (by a dataset). In both cases the model is replaced by fictional streams representing observed quantiles.

In future research, we want to focus on refining the validity of models by identifying the causes of process variability. The dynamic separation of inherent process variability from definable (identifiable, systematic, removable, special) variability will be essential for the reliability of the models. For this separation, working with Big Data processing will be very advantageous.

## 6. Funding

The research was funded by the Technology Agency of the Czech Republic, project number TAČR FW03010296 "Practical use of big data for an intelligent production flow decision system," principal investigator Jan Štindl (dataPartner s.r.o). This funding source financially supported only the first three authors.

A significant part of the paper, namely most of formulas in Section 3, was derived by the fourth author in his Bachelor's thesis (Šťovíček, 2023).

## References

Chen, D. Z. and Wang, H. (2013). Approximating points by a piecewise linear function. *Algorithmica*, 66:682–713.

Emir, K., Kruml, D., Paseka, J., and Selingerová, I. (2019). A simplified probabilistic validation of production flows.

In *Proc. of the 18th International Conference on Modeling and Applied Simulation*, pages 166–173. CAL-TEK SRL.

Emir, K., Kruml, D., Paseka, J., and Selingerová, I. (2021). Quantile based summation of random variables. In *Joint Proceedings of the 19th World Congress of the International Fuzzy Systems Association (IFSA), the 12th Conference of the European Society for Fuzzy Logic and Technology (EUSFLAT), and the 11th International Summer School on Aggregation Operators (AGOP)*, pages 523–529. Atlantis Press.

Kotler, P., Keller, K., and Chernev, A. (2021). *Marketing Management*. Pearson, Harlow Essex.

Kruml, D. and Paseka, J. (2018). Categoric simulation of production flows. In *Proc. of the 17th International Conference on Modeling and Applied Simulation*, pages 68–75. CAL-TEK SRL.

Liker, J. (2020). *The Toyota Way: 14 Management Principles from the World's Greatest Manufacturer*. McGraw Hill Education, New York.

Luo, G., Yi, K., Cheng, S.-W., Li, Z., Fan, W., He, C., and Mu, Y. (2015). Piecewise linear approximation of streaming time series data with max-error guarantees. *2015 IEEE 31st International Conference on Data Engineering*, pages 173–184.

McCullough, D. and Wade, E. (2003). Recursive enumeration of Pythagorean triples. *The College Mathematics Journal*, 34.

Porter, M. E. (1985). *The Competitive Advantage: Creating and Sustaining Superior Performance*. Free Press, New York.

Pottmann, H., Krasauskas, R., Hamann, B., Joy, K., and Seibold, W. (2000). On piecewise linear approximation of quadratic functions. *Journal for Geometry and Graphics*, 4:31–53.

Rother, M. and Shook, J. (2003). *Learning to See: Value-Stream Mapping to Create Value and Eliminate Muda*. Lean Enterprise Institute,US, Cambridge.

Šťovíček, J. (2023). *Use of Pythagorean triples for linear approximation (in Czech)*. Bachelor's thesis, Masaryk University, Faculty of Science, Department of Mathematics and Statistics, Brno. https://is.muni.cz/th/mfmfm/.