DOI: https://doi.org/10.46354/i3m.2019.mas.021

A SIMPLIFIED PROBABILISTIC VALIDATION OF PRODUCTION FLOWS

Kadir Emir[(a)], David Kruml[(b)], Jan Paseka[(c)], Iveta Selingerová[(d)]

[(a),(b),(c),(d)]Department of Mathematics and Statistics, Masaryk University
Kotlářská 2, 611 37 Brno, Czech Republic

[(a)]emir@math.muni.cz,[(b)]kruml@math.muni.cz,[(c)]paseka@math.muni.cz,[(d)]selingerova@math.muni.cz

ABSTRACT

This paper extends the stack validation algorithm in a probabilistic way. In other words, we introduce a new model for stack validation when the production parameters are random variables and the result is compared with a confidence interval. The major outcome of this simplified probabilistic model is to determine random variables merely by mean, variance, and skewness. This straightforwardly enables some direct, fast and consistent calculations by using certain properties of these moments.

Keywords: categoric simulation of production flow, signal, stochastic processing times, probabilistic stack validation

1. INTRODUCTION

In the industrial factory planning practice; planning capacity, material in volatile or changing demand situations often require high capacities and also cause material plan instability for both suppliers and factories. The decisions of the planner can be based on many factors such as period machine capacity, profit margins, holding costs, etc. One of the biggest challenges of any planner is dealing with the level of capacity (input, output, or stack) during any step of the production flow. At this point, standard and traditional deterministic methods are of little use. However, probabilistic and statistical methods come into play for predicting the underlying risk of capacity and material in planning, in advance.

The category theoretical description and simulation of production flow is developed in Kruml and Paseka (2018) by introducing the *signal model*. This model is perfectly suitable for modeling the production in an algebraic point of view, which is mass and regular on processes. But these processes do not have to be synchronized. That is, such processes can operate under different speeds and batch sizes, and we need high capacity stacks to eliminate such differences. Every stack is filled by outputs of preceding processes and drawn by inputs of the next processes. We refer *signals* for these time functions. In regular mass production, signals can be effectively encoded by *formal words* where we write x for a produced unit, o for a time unit, and powers to mean lengths and repetitions. For instance, the word $(o^5x)^{10}o^{30}$ stands for the signal A (Fig. 2) where production cycle o^5x comprises cycle time 5 and 1 produced unit, is repeated 10 times, and followed by the inactivity of time 30. Expansion of the word draws a walk in the mass-time, where x is a move in mass and o is a move in time.

Such a decomposition and composition of signals is a part of general principles formulated for all aspects of the production flow and formalized by terms of category theory, see Coecke and Paquette (2011). Moreover, we know that a production flow can be modeled within three modes – time, space (network), and mass (material and products). The composition principle enables us to figure out the flow as a tree-organized collection of subordinated jobs and solve them independently, simultaneously, and with predefined accuracy.

However, the crucial problem in our methodology is the behavior of the stack. The flow is always considered to be valid regarding the stack, in the given simulation interval, and the stock needs to be kept balanced within the lower and upper bounds of the stack. The stock function is defined as a sum of all input and output signals. Its evolution could look chaotic and computationally hard for a complete determination. Instead of this, the *stack validation algorithm* (SVA) is developed based on tree organization of signals. The algorithm starts with rough approximations on large time intervals and continues with detailed approximations only on smaller parts where the rougher methods fail to decide the stack validity. In other words, it uses the recursive search of critical moments and avoids detailed inspection elsewhere that makes it *lazy* and *effective*.

2. STATEMENT OF THE PROBLEM

In this paper, we skip the process of recursion and mainly focus on the inner step of SVA. Here, we have the signals (Fig. 1) with lower and upper linear approximations, and we also use their sums as approximations of the sum signal. This follows a discussion of whether the approximations overlap the bounds. We extend this idea to the probabilistic level and describe an analogous step of *probabilistic stack validation algorithm* (PSVA). The recursive part of the algorithm remains similar to that of SVA.

Afterward, we propose to use our signal model to incorporate the probabilistic and statistical issues into the planning process. To our knowledge, no similar formulation exists in the literature, which is directly applicable to the planning situation discussed above.

Such modeling situation results in stochastic process-

ing times, and it is the typical one where queueing effects Manitz (2008) or stochastic-flow network models Fiondella, Lin and Chang (2015) occur. Nevertheless, we can still use the regularity of processes.

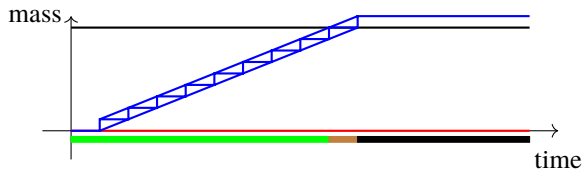

Figure 1: Stack validation

The lines enveloping the signal provide upper and lower approximations, which are much easier for calculations. Crossings with the upper bound (black) and the lower bound (red) of a stack indicate states of SVA (here we assume that there is no other signal affecting the stack). The green interval is surely correct, the black interval is surely incorrect (overfill), and the brown interval will be inspected with higher accuracy in a further recursion step.

We also expect that our probabilistic model could be useful also for any other composed probability systems including job production or project management – e.g., the network analysis technique PERT (see Pohl and Chapman (1987); Heagney (2016)).

This paper addresses the issues given above and makes the following contributions:

1. We extend the signal model from Kruml and Paseka (2018) to a probabilistic version without substantial changes.

2. Our approach can check the behavior of the whole system at each stack separately and hence to provide validity tests in a fast and reliable manner.

3. The method does not use the experimental approach (Monte Carlo simulation). Predictions are made by direct calculations based on fast transformations between probabilistic moments and quantiles.

3. STOCHASTIC SIGNALS

A geometric interpretation of the production flow is introduced in Kruml and Paseka (2018) through a certain surface in a space-time-mass. The model describes an ideal type of production with no errors or uncertainty. Here, we would like to describe a probabilistic extension of that model in the sense that any flow parameters could be considered as random variables. In this case, the resulting flow should not be imagined as a surface anymore but as a "fuzzy cloud" inside the space-time-mass – i.e., representing the density of possible runs of a given plan. We know that random effects are often dependent. For instance, an error of one process can affect the behavior of other processes or waste in series due to inappropriate machine setting. However, we suppose that many types of dependencies vanish when the random effects are resolved to elementary ones and positioned to appropriate segments in the model hierarchy.

Among other possibilities, the flow can be validated by the inspection of stacks. Namely, we test whether a stack keeps acceptable storage of produced units all the time, so it is never overfilled or lacked concerning defined bounds. We call this a *signal model* because the stack validation algorithm evaluates the input and output *signals*, i.e., time functions representing mass processed in time. The signal can be defined as a *formal word* and has a natural breakdown structure of subwords or subsignals. The main advantage of the model is that mass production is highly regular and repetitive. Moreover, the repetitions can be effectively encoded as *powers* in words. We claim that a large class of random effects can be modeled merely by adjusting these powers (Fig. 3, Fig. 4 and Fig. 5). The idea is quite general and also could be useful in planning or simulation of less regular applications, e.g., job production or project management.

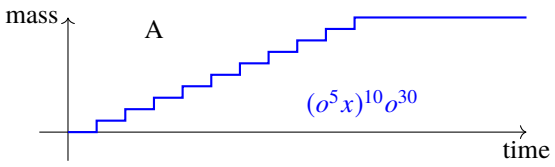

Figure 2: Signal A

The letters x, o mean mass (product) and time (delay) units, respectively. Signal A models a machine working in a regular regime that produces 1 mass unit per 5 time units. It totally produces 10 units and stops for 30 units.

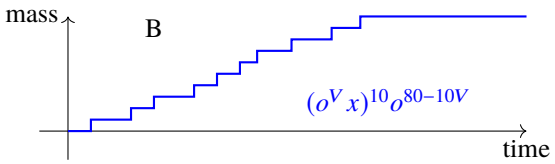

Figure 3: Signal B

Signal B is affected by irregularity in cycle time which is defined as random variable V.

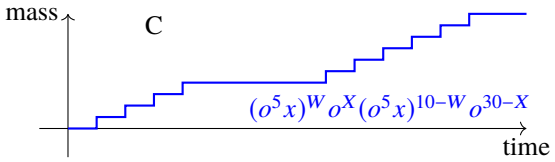

Figure 4: Signal C

In signal C, the machine works again in regular cycle time but the season is randomly divided to two parts by some stoppage. Here, we have two random variables: W represents the number of repetitions before the stoppage, and X is the time of the stoppage.

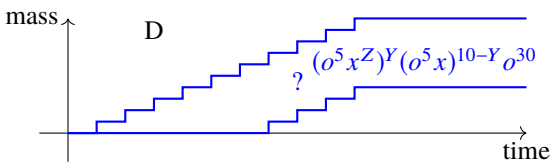

Figure 5: Signal D

Finally, signal D represents a mixture of two scenarios: in the first one the machine works well, in the second one the season starts with Y repetitions producing waste (and no products x). A scenario is selected by a random variable Z with Bernoulli distribution (see Ross (2014)) – the first scenario is performed if Z takes value 1, the second one if Z takes 0. All examples can be combined.

4. PROBABILISTIC STACK VALIDATION

The stack validation algorithm (SVA) respects the breakdown tree hierarchy of signals and iteratively search only on intervals where rough approximations do not provide decisive answers. This approach makes it more effective than straightforward simulation methods, e.g., discrete-time simulation. When we like to extend the algorithm to the probabilistic environment, we must reformulate the statement of the problem as well as to develop a fast calculus for manipulations with random variables.

First of all, *we give up any attempt to predict "global probability" that a given plan is capable*. This is a quite hard aggregation problem where all dependencies, hidden decisions, objective and subjective preferences should come into account. Instead of this, we still assume that the *plan evolves by small improvements made on particular processes, jobs, and orders.* The system communicates with the planner by predicting whether affected stacks perform well until the performance is considered satisfactory. But such a conclusion is completely under the responsibility of the planner. We must emphasize this fact because the probabilistic results are hardly 100% decisive and the planner should anticipate results like "the stack will be overfilled with a probability smaller than 5%".

According to the philosophy that "makes the planner responsible for everything", we let him/her set these *confidence intervals* for each stack separately and provide the validation concerning them. This can reflect thinking in a sense "*Overfill of that stack would mean a small temporal disorder next to the cutter but this happens often, and nobody bothers*" or "*Shortage of that stack is unpleasant, but I have some extra reserves of material to supply*".

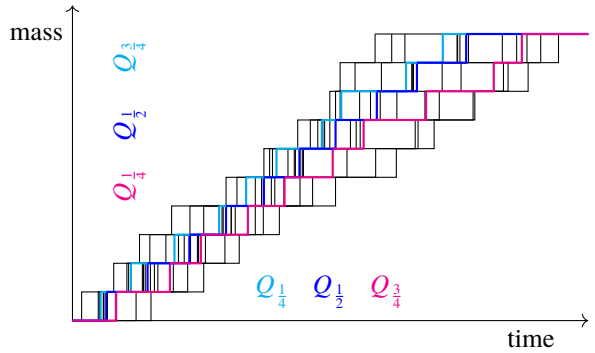

Figure 6: "Fuzzy cloud" representing a density of possible runs of a given plan

The "cloud" in Fig. 6 is drawn by multiple runs of a probabilistic process. The first, second, and third quartile is depicted by a cyan, blue, and magenta lines, respectively (From the mass point of view, the order is reversed).

Our main goal is to find *reliable approximations of the flow* such that we would certainly know that several possible runs out of such range are less than a given percentage. In other words, *we compare bounds of the stack with given quantiles of (random and time-varying) stock.* Since we consider only one-way transitions in the production network (two-ways transitions can be separated into two channels), the mass development of any signal has an obvious orientation (positive on inputs, negative on outputs). Of course, time only moves in a positive direction. Thanks to this, we find obvious but important fact about the flow: For a given *state* (t, N) (i.e., the event that Nth unit passed in time t) there is the same number of runs "on the right" (those which reach N after time t) and of runs "below" (those which reach N before time t). One can consider the flow as a collection of random walks or an analog of two-dimensional distribution. In this formalism, the above fact states that the marginal distributions along time and the marginal distributions along mass *share quantiles*, i.e., distribution of mass in time t reach in value N the same quantile like distribution of time in mass N in value t (Fig. 6 and Fig. 8). (To be more precise, for a quantile Q_p we should rather speak about the complementary quantile Q_{1-p} because of the orientation of the time and mass axes.)

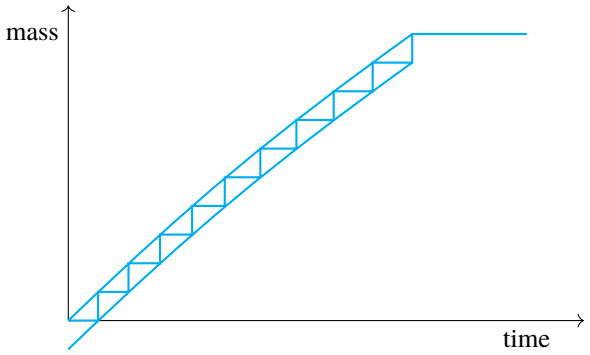

Figure 7: Upper and lower approximations of a quartile

By theory, the quartiles evolve as sequences linearly depending on n and $\sqrt{n}$ where n is the number of repetitions. Thus, they have upper and lower approximations by certain quadratic curves. These can be interpreted as quartiles of a *continuous random process*.

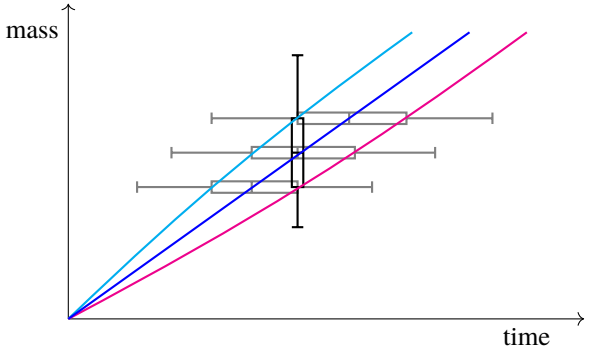

Figure 8: Swap of axes (from time to mass)

In Fig. 8 quartiles for fixed mass (gray) yield quartiles for fixed time (black). Consequently, we use them to estimate the vertical distribution.

Of course, this is not the case of densities – they are partial derivations of the cumulative density functions,

and there is no reason to be equal. Moreover, the two orthogonal distributions may have quite different shape properties. For example, if the signal is random in cycle time and we assume that all repetitions have the same distribution and are independent, then (after quite a small number of repetitions) the sum of times tends to a normal distribution, regardless of the type of distribution on one cycle (central limit theorem). On the other hand, the corresponding orthogonal distribution need not be symmetrical even in a large number of repetitions – we can see that there is more pointed density for small N and less pointed for large N because later units are produced with higher dispersion in time. Thus, the *skewness* of the distribution will be positive.

The described swap of axes enables us to study the mass (vertical) distribution using quantiles of time (horizontal) distributions. No matter which distributions enter the validation and how they are combined, we use the idea at least for the end of each iteration step in the SVA because we compare the mass distribution with the bounds of the stack. But it is also important for parallel summation of more than one random signals, which is a typical situation occurring at stacks.

The swap is known for certain classes of distributions. Namely, Poisson distribution is, in this sense, inverse to exponential distribution and normal distribution to a certain type of generalized inverse Gauss distribution. Our method is not precise, but it is simple and general.

5. L-ESTIMATORS AND MOMENTS

In practice, the random variables presented in the flow model are obtained from statistical analysis of real production datasets or guesses resulting from long-term experience. In many cases, the planner is not an expert neither on probability theory nor on mathematical statistics. Therefore he/she will be only able to use simple methods for the probability distribution.

A popular method is the 3-point estimation where just *minimum*, *middle*, and *maximum* values of the distribution are considered. The middle value is interpreted as a location parameter depending on context or planner's preferences; it can be mean, median or mode of the distribution. The minimum and maximum points determine variability. Since the minimum and maximum are very inclinable to chance, they can be replaced by much more robust quartiles Q_1 and Q_3, namely the quantiles of 25% and 75%. When we adopt the median Q_2 for the role of middle value, we model the distribution by triple quartiles (Q_1, Q_2, Q_3), that is, the values obtained by dividing the ordered dataset to four quarters. Similar approaches are examined in various works such as Bland (2015); Wan, Wang, Liu and Tong (2014) and also Hozo, Djulbegovic and Hozo (2005).

Important characteristics of the distribution can be guessed by the appropriate *L-estimators*. Namely, we will use the following three ones:

- $\frac{Q_1 + Q_3}{2}$ ... *midhinge*, location measure, replacement for mean,
- $Q_3 - Q_1$ ... *interquartile range*, scale or dispersion measure, replacement for standard deviation,
- $\frac{Q_1 + Q_3}{2} - Q_2$ ... difference of midhinge and median, *skewness measure*.

Such simplifications could be very inaccurate in many situations, e.g., when the modeled distribution is discrete or of exceptional shape. (A deterrent example is the Bernoulli distribution Z from signal D in Fig. 5.) Nevertheless, our experiments with many natural types of continuous unimodal distribution demonstrated that the L-estimators are in a surprisingly good coincidence with moments, and they provide an elementary calculus for summation of random variables. Notice that the transformation from quartiles to L-estimators is reversible. Hence one can also recover the quartiles from the L-estimators. (This is also the reason why we work only with a triple of quartiles instead of popular *five-point* or *seven-point* cases. Here the transformation would be more accurate but not reversible.) Thanks to this, we can also invert the procedure for special distributions including the discrete ones – we first find the moments by precise mathematics *and then* get *virtual quartiles* by the inverse transformation. The remaining step converting moments to the L-estimators is a linear adjustment of each of the moments and can be done concerning an assumed kind of distribution. That is, one can create a "dictionary" of frequently used distributions with specified values of the coefficients. In this manner, we can handle any distribution in two equivalent ways – either by the three quartiles (regardless they have the correct meaning, or they are just "ghosts") or by three moments, namely mean, variance, and skewness. The former triple is needed for the swap of axes, the later for a quick and stable summation of random variables which is based on the well-known *additivity* of the three moments (see Renyi (2007)):

Theorem 1 *Let X_1, X_2 be two independent random variables with the triples $(\mu_1, \sigma_1^2, \tau_1), (\mu_2, \sigma_2^2, \tau_2)$ of moments, i.e., mean, variance, and third central moment. Let $X = X_1 + X_2$. Then*

$$\mu = \mu_1 + \mu_2, \qquad \sigma^2 = \sigma_1^2 + \sigma_2^2, \qquad \tau = \tau_1 + \tau_2$$

are moments of the random variable X.

In particular, let $X_i, i = 1, \dots n$ be independent identically distributed random variables, and define:

$$X = \sum_{i=1}^{n} X_i.$$

If the fixed triple $(\mu_1, \sigma_1^2, \tau_1)$ will be the moment triple of any X_i, then

$$\mu = n\mu_1, \qquad \sigma^2 = n\sigma_1^2, \qquad \tau = n\tau_1$$

are moments of the random variable X.

For a random variable X with a moment triple (μ, σ^2, τ) we define $\gamma = \tau/\sigma^2$. We can think of γ as a parameter for *skewness*. If X_1, X_2 are independent random variables with moments triples $(\mu_1, \sigma_1^2, \tau_1), (\mu_2, \sigma_2^2, \tau_2)$,

$\gamma_1 = \tau_1/\sigma_1^2$, $\gamma_2 = \tau_2/\sigma_2^2$ and $X = X_1 + X_2$ then we obtain the parameter γ for X as

$$\gamma = \frac{\gamma_1\sigma_1^2 + \gamma_2\sigma_2^2}{\sigma_1^2 + \sigma_2^2}.$$

The midhinge is a direct guess of the mean and thus we will also denote it by $\widehat{\mu}$. The interquartile range is assumed to be a multiple of the standard deviation, i.e., the square root of the variance. To make the use more intuitive, we will consider half of the interquartile range and denote it by $\widehat{\sigma}$. Finally, the difference between midhinge and median will be denoted by $\widehat{\gamma}$. The parameter γ is assumed to be proportional to the scale in the sense that third central moment τ linearly depends on $\gamma\sigma^2$. This property yields direct formulas for the moments. Hence for a prediction of a sum of two independent random variables we will use the following rules:

$$\widehat{\mu} = \widehat{\mu}_1 + \widehat{\mu}_2, \quad \widehat{\sigma} = \sqrt{\widehat{\sigma}_1^2 + \widehat{\sigma}_2^2}, \quad \widehat{\gamma} = \frac{\widehat{\gamma}_1\widehat{\sigma}_1^2 + \widehat{\gamma}_2\widehat{\sigma}_2^2}{\widehat{\sigma}_1^2 + \widehat{\sigma}_2^2}. \ (*)$$

Consequently, the sum of n independent identically distributed random variables will be predicted as follows:

$$\widehat{\mu} = n\widehat{\mu}_1, \qquad \widehat{\sigma} = \sqrt{n}\widehat{\sigma}_1, \qquad \widehat{\gamma} = \widehat{\gamma}_1.$$

Let us recall that

$$\widehat{\mu} = \frac{Q_1 + Q_3}{2}, \quad \widehat{\sigma} = \frac{Q_3 - Q_1}{2}, \quad \widehat{\gamma} = \frac{Q_1 + Q_3}{2} - Q_2,$$

and derive the inverse transform

$$Q_1 = \widehat{\mu} - \widehat{\sigma}, \qquad Q_2 = \widehat{\mu} - \widehat{\gamma}, \qquad Q_3 = \widehat{\mu} + \widehat{\sigma}.$$

6. EXAMPLE

To demonstrate the model we consider a stack with the input signal $s_1 = o^{20}(xo^Y)^{120}$ and the output signal $s_2 = o^{25}(x^6o^Z)^{20}$ where Y, Z both have triangular distributions with min/max/mode values $0/1/0.5$ or $0/6/3$, respectively. This yields quartiles

$$Q_{11} = \frac{1}{2\sqrt{2}} \approx 0.354, \qquad Q_{12} = \frac{1}{2},$$
$$Q_{13} = 1 - \frac{1}{2\sqrt{2}} \approx 0.656, \quad Q_{21} = \frac{3}{\sqrt{2}} \approx 2.121,$$
$$Q_{22} = 3, \qquad Q_{23} = 6 - \frac{3}{\sqrt{2}} \approx 3.879,$$

and L-estimators

$$\hat{\mu}_1 = 0.5, \qquad \hat{\sigma}_1 = 0.156, \qquad \hat{\gamma}_1 = 0,$$
$$\hat{\mu}_2 = 3, \qquad \hat{\sigma}_2 = 0.879, \qquad \hat{\gamma}_2 = 0.$$

Let us look at the finishing time of s_1. The model predicts that the quartiles evolve to

$$Q_{11,13} = 20 + 120\hat{\mu}_1 \mp \sqrt{120}\hat{\sigma}_1 \approx 20 + 60 \mp 2.236$$
$$= 77.764, 82.236,$$
$$Q_{12} = 20 + 120\hat{\mu}_1 - \hat{\gamma}_1 = 20 + 60 + 0 = 80.$$

Five simulations, each with 1000 runs, provided datasets with:

experiment	Q_{11}	Q_{12}	Q_{13}
1	78.445	79.916	81.545
2	78.464	79.911	81.222
3	78.436	80.086	81.938
4	78.468	79.993	81.689
5	78.537	79.944	81.511

The result has a significantly different interquartile range (or standard deviation). This difference is caused by the fact that one cycle distribution is assumed to be triangular while its 120th power is almost normal. But ratio $\hat{\sigma}/\sigma$ is about 0.717 for triangular distribution while 0.675 for normal distribution. When the difference is multiplied by $\sqrt{120}$, we get an error 0.470 which agrees with the difference between modeled values and experimental results. (We can conclude from this that the input quartiles of cycle times should be rather obtained from repeated cycles or large datasets than from a single cycle and triangular expectation.)

Now let us look on stock at time 40 for both signals and calculate quartiles in the mass axis. The quartile guesses appear as solution of equations, cf. Fig. 8:

$$\hat{\mu}_i\frac{Q_{i1}}{n_i} + \hat{\sigma}_i\sqrt{\frac{Q_{i1}}{n_i}} = 40 - t_i,$$
$$\hat{\mu}_i\frac{Q_{i2}}{n_i} - \hat{\gamma}_i = 40 - t_i,$$
$$\hat{\mu}_i\frac{Q_{i1}}{n_i} - \hat{\sigma}_i\sqrt{\frac{Q_{i1}}{n_i}} = 40 - t_i.$$

where $i = 1, 2$ is an index of signal, $n_1 = 1, n_2 = 6$ the batch sizes, and $t_1 = 20, t_2 = 25$ the start times.

This yields

$$Q_{11} \approx 38.190, \qquad Q_{12} = 40, \qquad Q_{13} \approx 41.896,$$
$$Q_{21} \approx 26.319, \qquad Q_{22} = 30, \qquad Q_{23} \approx 34.195,$$

hence

$$\hat{\mu}_1 \approx 40.043, \qquad \hat{\sigma}_1 \approx 1.853, \qquad \hat{\gamma}_1 \approx 0.043,$$
$$\hat{\mu}_2 \approx 30.257, \qquad \hat{\sigma}_2 \approx 3.938, \qquad \hat{\gamma}_2 \approx 0.257.$$

For the subtraction signal $s = s_1 - s_2$ we get from (*)

$$\hat{\mu} = 9.786, \qquad \hat{\sigma} \approx 4.352, \qquad \hat{\gamma} \approx -0.203,$$

and by transforming it back

$$Q_1 \approx 5.434, \qquad Q_2 \approx 9.989, \qquad Q_3 \approx 14.138.$$

Notice now that PSVA still works with envelopes of

signals where the quartile guesses are shifted up and down by constants depending on the stair step (batch size) of each of signals. Here the upper guess should be shifted by 1 (batch of positive s_1) and the lower guess by 6 (batch of negative s_2).

Again, from experiments of 1000 runs we get datasets

experiment	Q_{11}	Q_{12}	Q_{13}
1	3	8	12
2	3	8	11
3	3	8	11
4	3	7	11
5	3	8	12

which are close to the centre of expected signal range, and display expected variance and skewness. (The mass distribution represents stored products, thus it is discrete with support on integers.)

The model provided an excellent performance also for other settings.

7. APPROXIMATING CURVES

Like in the ideal production SVA, we need to determine intervals where stock is in given bounds and where it is not. The PSVA assumes that the enveloping approximations of a certain quantile are always quadratic curves of form $at + b\sqrt{t + c} + d$. In the simplest case of one signal of a regularly working process, this has an obvious interpretation: a is mean production speed, b, c are related to standard deviation, b is positive for Q_3 and negative for Q_1, and d comprises actual stock at the beginning of the tested time interval and eventually the skew component γ. Since the Q_1 and Q_3 differ only on the sign of b, the function $(Q_3 - Q_1)^2/4$ (an L-estimator corresponding to variance) is only a linear function of time. Summing of such signals would be very simple because we could use the formulas (*) to calculate the coefficients a, b, c, d.

But in general situations, there is no relation between the values b for Q_1 and Q_3. For example, when the process start is random, part of runs already draws the parabolic shape of Q_3, i.e., the parameter b is non-zero, while the other runs still "do nothing" and the function Q_1 is constant with $b = 0$ on the initial interval of production. Evolution of L-estimators is non-linear, and summed signals could have complicated formulas for their exact values.

However, we still consider that the L-estimators are replacing mean, variance, and the third central moment and that they evolve linearly. This yields formulas

$$\begin{aligned}\hat{\mu} &= (1-t)\hat{\mu}_0 + t\hat{\mu}_1,\\ \hat{\sigma}^2 &= (1-t)\hat{\sigma}_0^2 + t\hat{\sigma}_1^2,\\ \hat{\gamma} &= \frac{(1-t)\hat{\gamma}_0\hat{\sigma}_0^2 + t\hat{\gamma}_1\hat{\sigma}_1^2}{(1-t)\hat{\sigma}_0^2 + t\hat{\sigma}_1^2}\end{aligned}$$

where the index 0 stands for begining of a time interval, index 1 for its end, and $t \in [0, 1]$. Thus we extrapolate the values of L-estimators to any time just from the values at some critical points. It is reasonable to consider all the points where some of the quartile curves are broken.

The assumption is correct for the central part of the process activity but not exact outside it, as the Fig. 9 shows.

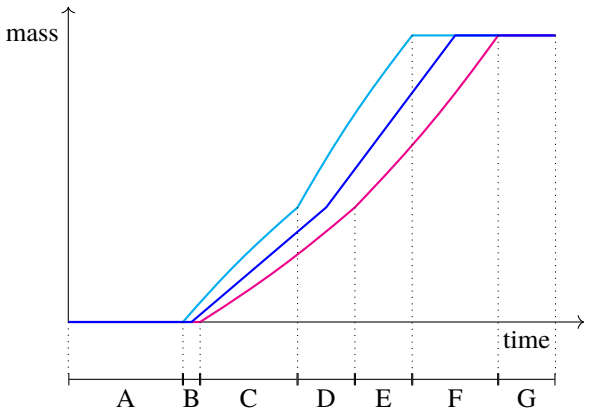

Figure 9: Evolution of L-estimators

In A, there is no activity. In B, the process may have started or not yet. In C, the process surely works in one regime. In D, the process still works in the first regime or perhaps it has switched to the second (faster) regime. In E, the process surely works in the second regime. In F, the process still works in the second regime, or it has already stopped. In G, the process has undoubtedly finished. The periods A, C, E, and G are parametrized exactly, the periods B, D, F are simplified. However, the error caused by this assumption is relatively small concerning the error caused by the simplified calculus of L-estimators.

Another objection to the model is that it does not assume that the process could start sooner than Q_1 or finish later than Q_3, but both cases may quickly happen. Such a problem can be fixed by adding two more parameters Q_0 and Q_4 according to Tukey (see Tukey (1977)), representing reasonable bounds of the distribution. They could be used in situations when $Q_1 = Q_2 = Q_3$, that is, when the distribution seems to be trivial (and the process to be deterministic) but we know that it is not. These Q_0 and Q_4 could be interpreted as "0th" and "4th" quartile and identified with $\hat{\mu} \pm 3\hat{\sigma}$. Then there are analogous transformation rules between quartiles and L-estimators and Q_0, and Q_4 hold the quadratic nature like Q_1 and Q_3.

Anyway, the proposed model approximates the probability density of a signal at all points by simply parametrized curves. Both quartiles and L-estimators have quadratic parametrization, and hence the same is true for sums of signals. The validity of a stack concerning a given confidence interval is then answered by solving an appropriate quadratic equation.

8. CONCLUSION

The described method may resemble a Fourier transform of functions. Summation of general random variables can be provided by discrete or continuous convolution. However, this can be very difficult or at least computationally hard. Instead of this, we parametrize the variables just by the three quartiles, convert them to certain L-estimators and work with those as with moments. Assuming the additivity, we quickly get L-estimators of the sum and convert them back to quartiles (see Fig. 10).

The quartile parametrization enables to swap between the time and mass axes easily. The only thing we need to do is to find the intersections of quartile curves with a line. Since mean evolves linearly and the skew parameter is constant for a repetitive process, we consider such parametrization for a general description of all approximations. Consequently, the crossing points can be found as solutions of certain linear or quadratic equations. The result is converted back to any of the forms and used for the next operation or to recover the resulting distribution.

Using the method, we can extend SVA to the probabilistic environment without substantial changes in its general principles and reasoning. The derived calculus is rough and simplified but still robust and highly effective for repetitive processes. It completely avoids the traditional Monte Carlo approach in simulation. Our experiments show that the proposed model is not only quick but also very realistic and consistent.

From another point of view, we know from Kruml and Paseka (2018) that the production flow without error or uncertainty forms a monoidal category structure with tensor product (Coecke and Paquette (2011)). In this paper, we now replace the certainty with uncertainty and some probabilistic errors. On the other hand, it is well known that there is a close relationship between fuzzy sets and probability theory; see Dubois, Nguyen and Prade (2000) for more details. Consequently, there might be again a monoidal category structure somehow related to fuzzy categories (Walker (2004)). In such a case, it is an interesting question how fuzzy categories and monoidal categories play a role together to model this structure in the light of category theory.

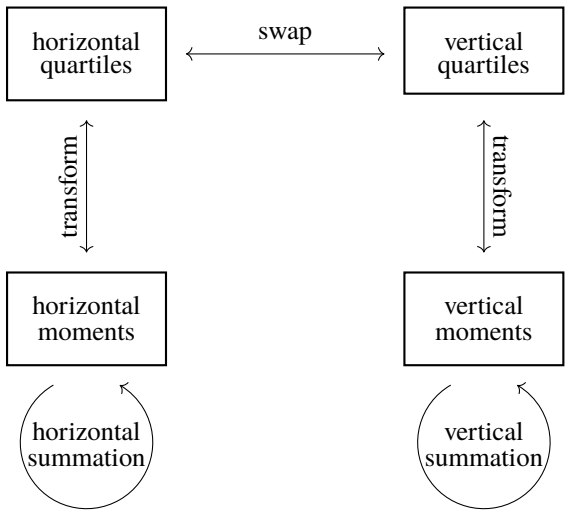

Figure 10: Computational strategy

ACKNOWLEDGEMENTS

We thank the anonymous referees for the extensive reading and contributions to improve our presentation of the paper.

Research of the first author was supported by the project Mathematical structures 8 (MUNI/A/1186/2018). Research of the second (corresponding) author was supported by the project New approaches to aggregation operators in analysis and processing of data, Nr. 18-06915S by Czech Grant Agency (GAČR). Research of the last two authors was supported by the project Research and Development of Information System for Value Flow Optimization and Production Planning financed by the Ministry of Industry and Trade: project CZ.01.1.02/0.0/0.0/15_019/0001370.

DECLARATION OF CONFLICTING INTERESTS

The author(s) declared no potential conflicts of interest concerning the research, authorship, and publication of this article.

REFERENCES

Bland M., 2015. Estimating mean and standard deviation from the sample size, three quartiles, minimum, and maximum. International Journal of Statistics in Medical Research 4 (1), 57–64.

Coecke B., Paquette É. O., 2011. Categories for the practicing physicist. In: Coecke B., ed. New Structures for Physics. Berlin: Springer, Lecture Notes in Physics, vol. 813, 173–286.

Dubois D., Nguyen H. T., Prade H., 2000. Possibility theory, probability and fuzzy sets. Misunderstandings, bridges and gaps. In: Fundamentals of fuzzy sets. Dordrecht: Kluwer Academic Publishers, 343–438.

Fiondella K., Lin Y., Chang P., 2015. System performance and reliability modeling of a stochastic-flow production network: A confidence-based approach. EEE Transactions on Systems, Man, and Cybernetics: Systems 45 (11), 1437–1447.

Heagney J., 2016. Fundamentals of project management. McGraw-Hill.

Hozo S. P., Djulbegovic B., Hozo I., 2005. Estimating the mean and variance from the median, range, and the size of a sample. BMC Medical Research Methodology, 5:13.

Kruml D., Paseka J., 2018. Categoric simulation of production flows. Proceedings of the 17th International Conference on Modeling and Applied Simulation, pp. 68–75. 17-19 September 2018, Budapest (Hungary).

Manitz M., 2008. Queueing-model based analysis of assembly lines with finite buffers and general service times. Computers & OR 35 (8), 2520–2536.

Pohl J., Chapman A., 1987. Probabilistic project management. Building and Environment 22 (3), 209–214.

Renyi A., 2007. Foundations of probability. Dover: Dover Publications, Dover Books on Mathematics.

Ross S.M., 2014. A first course in probability. 9th ed., international ed. Boston, MA: Pearson.

Tukey J.W., 1977. Exploratory Data Analysis. Boston, MA: Pearson.

Walker C.L., 2004. Categories of fuzzy sets. Soft Computing 8 (4), 299–304.

Wan X., Wang W., Liu J., Tong T., 2014. Estimating the sample mean and standard deviation from the sample size, median, range and/or interquartile range. BMC Medical Research Methodology 14 (1), DOI: 10.1186/1471-2288-14-135.

AUTHORS BIOGRAPHY

KADIR EMIR

is a Ph.D. student at the Department of Mathematics and Statistics of Masaryk University in Brno, Czech Republic; under the main supervision of Professor Jan Paseka. He received his first Ph.D. in 2016 in the field of algebraic topology. His current research area spans the category theoretic approach to quantum theory and production networks.

DAVID KRUML

is an assistant professor at the Department of Mathematics and Statistics of Masaryk University in Brno, Czech Republic. He received his Ph.D. in 2003. He is primarily interested in algebraic quantum structures, category theory, and game theory. He currently works as a key person on the project "Research and Development of Information System for Value Flow Optimization and Production Planning" financed by Ministry of Industry and Trade: project CZ.01.1.02/0.0/0.0/15_019/0001370 (2017–2019).

JAN PASEKA

is an associate professor at the Department of Mathematics and Statistics of Masaryk University in Brno, Czech Republic. He received his Ph.D. in 1990. His professional field of interests includes a non-commutative generalization of topology, quantum, and many-valued structures. He was a principal investigator of several research projects. He currently supervises at Masaryk university the project "Research and Development of Information System for Value Flow Optimization and Production Planning" financed by Ministry of Industry and Trade: project CZ.01.1.02/0.0/0.0/15_019/0001370 (2017–2019).

IVETA SELINGEROVÁ

is a researcher at Masaryk Memorial Cancer Institute and lecturer at the Department of Mathematics and Statistics of Masaryk University in Brno, Czech Republic. She received her Ph.D. in 2015. Her research activities include biostatistics, survival analysis, clinical trials, kernel estimation, or numerical methods. She worked within the project "Research and Development of Information System for Value Flow Optimization and Production Planning" financed by Ministry of Industry and Trade: project CZ.01.1.02/0.0/0.0/15_019/0001370 (2017–2019) on modeling and testing for simulation of production flows with a stochastic character.