# Simulation – A Support for Controllers Decision Process

*Veit Köppen*[1]
*Annette Hausmann*[1],
*Hans-J. Lenz*[1]

**Abstract:**

*Operative model-based Controlling checks the integrity of business figures, i.e. correctness with respect to reality. This data results from the business processes managed by various departments. The figures are represented as random variables and related by a non-linear system of equations. The random variables in each equation are connected by dyadic arithmetic operators like multiplication and addition. Business figures are normally not crisp, but random due to measurement errors. Consequently, computation of the corresponding probability distributions is usually not a trivial task.*
*In this paper a probability distribution function type is proposed which is not limited to a Gaussian distribution; any computable target probability function can be used. As there does not exist a parametric class of probability functions which is closed under folding by an arithmetic operation, MC-simulation is used. In this paper we study the Metropolis-Hastings (MH) method. Other Markov Chain Monte Carlo (MCMC) methods are not equally effective and easily implemented. Starting point of the simulation is the distribution of the random variables corresponding to some given data sets. The business figures of interest are simulated from a so-called target function. The distribution of interest can be estimated by simulation from the target function. The results may differ due to different target functions. The efficiency of different functions is examined. From a simulated data set any moment of interest like the mean or standard deviation or higher moments can be computed. Different distributions and their moments are analyzed. We are interested in efficient estimation. Increasing the variances of distributions leads to imprecise results. Therefore adaptive statistical methods are applied. Also a change of the distribution type may lead to improper simulations. We found out, that the MH method must be adapted for functions on a finite interval. Imprecision due to sampling and estimation must be kept as small as possible. If the target function has a large variance, the simulation method tends to become unreliable. Outliers may show-up. Consequently, robust estimation becomes mandatory. The kind of trimming is not too much influential, but more data has to be simulated in order to compensate the loss of data. There exists a trade-of between a computational slow-down and improved results.*
*The great advantage of using the Metropolis-Hastings algorithm is that irrespective from the input values, the simulated data follows exactly the function of interest. In addition to that, the estimators for business figures are very precise. The given data can be tested against the simulated data. A hypothesis test can be run, enabling the controller to either accept or reject the data. The power of the hypothesis test is sufficient given an $\alpha$-level. This implies that the risk of a wrong decision in case of non-consistent data is small. So simulation supports operative controlling based on models and contributes to improve the precision of business figures.*

## Keywords:

Business figures, Monte Carlo Methods, Metropolis-Hastings Algorithm, Simulation

---

[1] Freie Universität Berlin, Germany.

## 1. Introduction

Business figures found in many business reports or in Balanced Score Cards describe companies in a very comprehensive way. Many decisions are supported by those figures. A vital question is: Can the management be sure, that the figures are correct?
Most of the business processes produce figures which are measured. Of course, figures may be corrupted by errors. So the business figures can be interpreted as random variables. The system of equations which is related to the figures is a non-linear system with dyadic arithmetic operators connecting the variables in each equation. We assume that these equations are separable, i.e. each variable in each equation can be uniquely separated from the remaining set of variables. A classical system of business figure is the DuPont-System, which will be investigated here. Other systems may differ in the equations, but can be handled as well, i.e. Business Score Cards.
Markov Chain Monte Carlo simulation is a helpful tool to investigate random variables even if it is not possible to calculate the distribution function. The Metropolis-Hastings algorithm can be used to generate the probability distribution of random variables. This method can be easily implemented for instance in SAS, SPlus or R, and also proves to have good performance.

## 2. The DuPont-System of Business Indicators

In 1919 the chemical "DuPont" developed a system of business figures that became known as "ROI-Tree". It has a structure with the return on investment as the root node. It consists of seven variables and four equations.

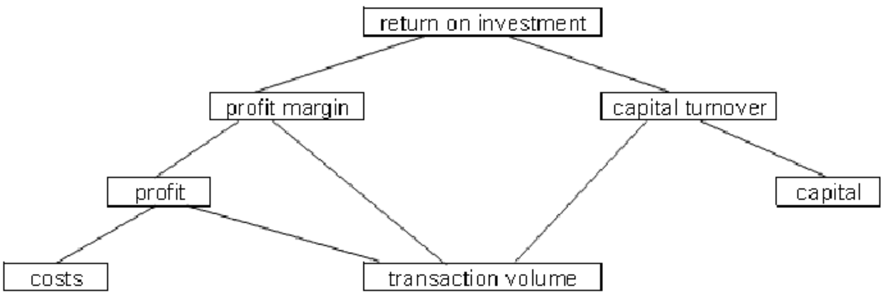



Figure 1: ROI-Graph of the "DuPont"-schema

The equation system is:

$$\begin{aligned} \textit{profit} &= \textit{transaction volume} - \textit{costs} \\ \textit{return on investment} &= \textit{profit} \,/\, \textit{capital} \\ \textit{profit margin} &= \textit{profit} \,/\, \textit{transaction volume} \\ \textit{capital turnover} &= \textit{transaction volume} \,/\, \textit{capital} \end{aligned}$$

Types of variables are:

- endogenous variables:
  - profit
  - return on investment
  - profit margin
  - capital turnover
- exogenous variables:

- transaction volume
- costs
- capital

In some applications some variable can only be estimated or measured with large precision. Other variables have a restricted range due to instructions or guidelines, e.g. the company's profit should be in a fixed interval. All of the given equations considered are separable, and that is why each variable can uniquely be isolated and form the left hand side of an equation. For instance, the profit can be calculated by

$$profit = profit\ \ margin\ \ *\ \ transaction\ \ volume\ .$$

## 3. Simulation

Due to measurement errors or data recording problems business figures are generally considered to be not crisp. Sampled data can be used to determine or specify distribution functions, but this slows down any computing performance. Distribution functions can also be specified by domain experts, but should be validated before being used. In the no data case one has to trust in the expert knowledge and prior information available.
Computation of the corresponding probability function is usually not a trivial task. We use Markov Chain Monte Carlo Methods. Therefore the distribution functions are not limited to the Gaussian distribution family. Because of interests of the flexibility and efficiency of the Metropolis-Hastings algorithm any distribution function can be used, cf. Hastings (1970), Chib (2004).

### 3.1. Extending the Metropolis-Hastings Algorithm

In the initialization phase the probability functions of the given variables are determined. Furthermore a proposal distribution is chosen for each exogenous variable. These distributions should need low computational efforts, because random samples have to be generated from that proposal distribution. To reduce the sampling costs the shape of proposal should be as close as possible to the desired probability function. If this is not possible, the sampling size has to be extended caused by the so called burn-in phase. The size of the sample depends upon two parameters are used: the number of particles for the Metropolis-Hastings algorithm and the number of repetition. The second parameter describes how many simulated means of particles per run will be used to estimate the distribution function of the exogenous variables.
In the second phase of this extended Metropolis-Hastings algorithm all exogenous variables are simulated. This is done is like in the Metropolis-Hastings algorithm.
In the third phase distributions of the endogenous variables are estimated using the equations of the corresponding sampled exogenous variables.
The number of iterations for these two phases is kept as parametric constant.
At the end a sample for all variables is given. So tests can be performed and given figures can be checked. In order to cut all the particles from the simulated data, the first means should not be a part of the analysis. The amount of data that has to be cut out depends on the proposal distribution.

The extended MH algorithm:

Fix the repetition size and the number of particles.
Exogenous variables are those with a known probability function.
Endogenous variables are described by the DuPont-system.

1. Initialize the given exogenous variables with proposal distributions and target probability functions.
2. Draw sample from the exogenous variables by using standard Metropolis-Hastings algorithm.
3. Derive the distribution of the endogenous variables from the equation-system.

4. Compute the means and variances of endogenous and exogenous variables and store them.
5. If the repetition size is not reached, go to 2.

The stored means and variances differ from run to run due to sampling errors.

### 3.2. Evaluation of the Algorithm

An evaluation of the extended MH algorithm requires to check whether or not the quality so simulation depends upon possible operators. The arithmetic operations are addition, subtraction, multiplication and division. Moreover different probability functions should also be considered: Gaussian distribution, triangle distribution and a Gaussian distribution with contamination.
The Gaussian distribution is frequently used in practice and research. It allows to use the mathematical calculus. The triangle distribution is also commonly used in practice and is a distribution function defined on a closed interval.
To check the results from the simulation the mean can be calculated or approximated by the following equations, cf. Moods et al (1974). The underlying assumption is that the random variables $X_i$ (i = 1,2) are identically and independently distributed. It follows for the means (expected values):

$$E[X_1 \pm X_2] = E[X_2] \pm E[X_2] \quad (1)$$

$$E[X_1 \cdot X_2] = E[X_1] \cdot E[X_2] \quad (2)$$

$$E[X_1 / X_2] \approx E[X_1] / E[X_2] + (E[X_1] \cdot Var[X_2]) / E^3[X_2] \quad (3)$$

The variances can be approximated by:

$$Var[X_1 \pm X_2] = Var[X_1] + Var[X_2] \quad (4)$$

$$Var[X_1 \cdot X_2] = E^2[X_1] \cdot Var[X_2] + E^2[X_2] \cdot Var[X_1] + Var[X_1] \cdot Var[X_2] \quad (5)$$

$$Var[X_1 / X_2] \approx (E[X_1] / E[X_2])^2 \cdot (Var[X_1] / E^2[X_1] + Var[X_2] / E^2[X_2]) \quad (6)$$

The following estimators are used for the mean and the variance of the simulated data:

Mean-Estimator: $\hat{\mu} = 1/T \sum_T X_i$ Variance-Estimator: $\hat{\sigma}^2 = 1/(T-1) \sum_T (X_i - \hat{\mu})^2$

In this case T describes the sampled size of the simulated data.

Now some examples will be discussed.

#### 3.2.1. Folding of two distributions under addition

As an example of addition we use

$$transaction\ \ volume = profit + costs$$

(i) Gaussian distribution
The exogenous random variables are profit and costs, and the endogenous random variable is transaction volume.
We specify the assumed Gaussian probability functions as follows:

$profit \sim N(20,1)$ and $costs \sim N(80,4^2)$.

To reduce simulation effort, the proposal distribution is the Gaussian distribution with the same parameters, too. This reduces the costs of simulation and does not alter the results of simulation.

By running the algorithm with 1000 particles (sampled values) per experiment and 5000 repetitions of experiments the estimated means and standard deviations of the exogenous variables do not differ from the desired levels. The simulated data for the transaction volume has an estimated mean $\hat{\mu} \approx 100$ and an estimated variance $\hat{\sigma}^2 \approx 17$. Using equation (1) to calculate the mean of the transaction volume leads to $\mu = 100$. Equation (4) leads to a theoretical variance $\sigma^2 = 17$.

As the exogenous variables are Gaussian distributed, the sum of two Gaussian random variables is also a Gaussian distribution. Thus a Kolmogorov-Smirnov goodness of fit test can be used based on the simulated data of the transaction volume. The $H_0$-hypothesis of all tests is that the simulated data is corresponding to a Gaussian distribution with mean 100 and variance 17. This test should be started directly after the Metropolis-Hastings phase. A 5000 times iteration follows. The p-values produced from each iteration are stored.

The mean of the p-values from the simulations delivers $\overline{p} \approx 50.4\%$. The variance of the p-Values of the Kolmogorov-Smirnov Goodness of fit tests is $\sigma_{\overline{p}}^2 \approx 29\%$. The minimum is 2.7% and the maximum is 100%. In this way it can be safely derived, that the simulation for a Gaussian probability function with the plus-operator is a good choice.

(ii) Triangle distribution

The next example uses the same equation with a triangle probability function. Profit is distributed in the interval [19, 21] with peak (mode) at 20. Costs are distributed in the interval [76, 84] with peak at 80. The proposal distributions are now a uniform distribution of profit in the interval [19, 21] and of costs in the interval [76, 84].

The estimators for the triangle distribution parameters are:

lower bound: $\hat{l} = 1/T \sum_T \min(X_i)$,

peak $\hat{p} = 1/T \sum_T \hat{\mu}_i$ and

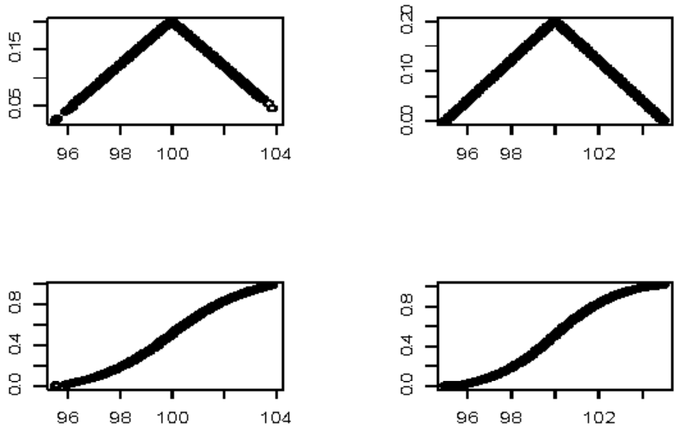
upper bound $\hat{u} = 1/T \sum_T \max(X_i)$.

The algorithm is again executed with the same simulation experiment parameters as in the case of (i). The estimated means and the lower and upper bounds of the exogenous variables do not differ much from the desired functions. The simulated lower bound $l$ has a mean $avg(\hat{l}) = 95.88$, but a minimum $\min(\hat{l}) = 95.12$. The simulated upper bound of the transaction volume has a mean $avg(\hat{u}) = 104.13$ and a maximum $\max(\hat{u}) = 104.81$. The mean of the simulated endogenous variable is $avg(\hat{p}) = 100$.

The theoretical results of the distribution should be in the interval of [95; 105]. So the minimum for the lower bound and the maximum for the upper bound are the better descriptors for the theoretical distribution. Testing of the simulated exogenous variables against the theoretical distribution the p-value of a two-sided Kolmogorov-Smirnov test has a value of about 0.

The p-value of the Kolmogorov-Smirnov test is derived as in the section of Gaussian distribution.

Figure 2 makes clear that the simulation in the regions of the upper and lower bounds are bad. The simulated density and distribution function are compared with their expected functions in figure 2.




simulated expected

**Figure 2: Simulated and expected triangle function distributions**

(iii) Contaminated Gaussian distribution
In our third example profit and costs are now corresponding to a 2-peak-Gaussian distribution. This probability function is described by:

$$(1-\varepsilon)\cdot N\left(\mu_1,\sigma_1^{\,2}\right)+\varepsilon\cdot N\left(\mu_2,\sigma_2^{\,2}\right) \text{ with } \varepsilon\in[0,1]. \qquad (7)$$

For this example, the profit distribution has the parameters:

$$\mu_1 = 21 \quad \sigma_1^{\,2} = 2 \quad \mu_2 = 16 \quad \sigma_2^{\,2} = 1 \text{ if } \varepsilon = 0.1$$

and the probability distribution of costs has the parameters:

$$\mu_1 = 75 \quad \sigma_1^{\,2} = 4 \quad \mu_2 = 90 \quad \sigma_2^{\,2} = 3 \text{ if } \varepsilon = 0.3\,.$$

Our proposal distribution is a Gaussian distribution and the starting values for the Metropolis-Hastings algorithm are uniformly distributed in the interval [10, 30] for profit and [70, 100] for costs.
The simulation again runs with 1000 particles and a repetition rate of 5000. The simulated mean of profit is $\hat{\mu}_{profit} = 20.52$ and the simulated mean of costs is $\hat{\mu}_{costs} = 79.10$. Using equation (1) leads to the value of $\mu = 99.62$. The simulated mean of transaction volume is $\hat{\mu} = 99.62$.
The standard deviation of profit is $\hat{\sigma}_{profit} = 2.51$ and of costs $\hat{\sigma}_{costs} = 7.82$. Equation (4) therefore leads to the theoretical value of $\sigma^2 = 67.45$ variance which is supported by simulated variance of $\hat{\sigma}^2 = 67.42$. The Kolmogorov-Smirnoff Goodness of fit test could not be applied to the data, because the distribution function of the result is not known and can not be derived. Nevertheless a simulated distribution of the equation is shown in figure 3. The 2-peak-Gausian distribution can very easily been detected in all exogenous variables, i.e. profit and costs.

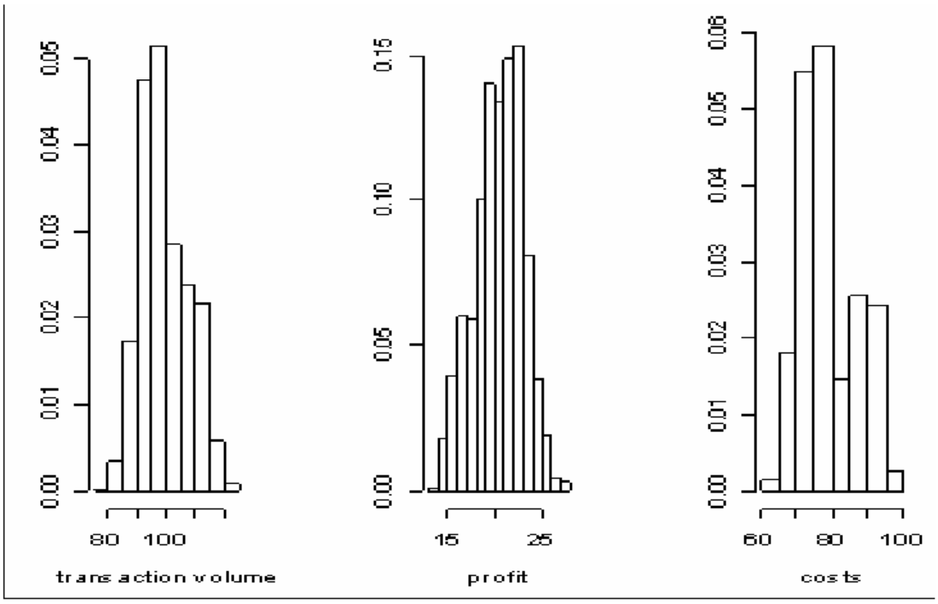



**Figure 3: Simulated distribution of transaction volume, profit and costs from a 2-peak Gaussian probability function**

For the operation "-" equivalent results can be derived. Therefore we skip this topic. As a first conclusion, simulation using the Metropolis-Hastings algorithm leads to good results not only in the endogenous variables, but also in the exogenous variables under all linear operations. Some problems can arise if the variance is very large, since the simulated data are affected. A proposal of tackling this problem is given in the last section.

### 3.2.2. Folding of two distributions under Multiplication

As an example for a non-linear operation the following equation is used:

$$profit = ROI * capital\ .$$

The exogenous variables are return on investment (ROI) and capital. Profit is the endogenous variable.

(i) Gaussian distribution

In the case of a Gaussian probability function the variables are distributed as:

$$ROI \sim N\left(0.25,\, 0.025^2\right) \quad \text{and} \quad capital \sim N\left(80, 4^2\right).$$

The proposal distribution is as in section 3.2.1 a Gaussian distribution with $\mu_{ROI} = 0.25$ and $\sigma^2_{ROI} = 0.025^2$, as well as $\mu_{capital} = 80$ and $\sigma^2_{capital} = 4^2$. The parameters of the simulation are 1000 particles and 5000 repetitions. The theoretical mean can be calculated by formula (2) and is $\mu = 20$. The simulated mean of profit is $avg(\hat{\mu}) = 19.99$. The variance which is computed by equation (5) and is equal to $\sigma^2 = 5.01$. The simulated variance is $\hat{\sigma}^2 = 4.99$. A Kolmogorov-Smirnov goodness of fit test for all simulated particles per run has a mean of $\overline{p} = 0$. The underlying distribution is a normal distribution with the theoretical mean and variance. Because of the zero values of these two-sided tests, the $H_0$-hypothesis that the simulated data is a Gaussian distribution must be rejected. In Figure 4 it becomes obvious that the simulated distribution is skew.

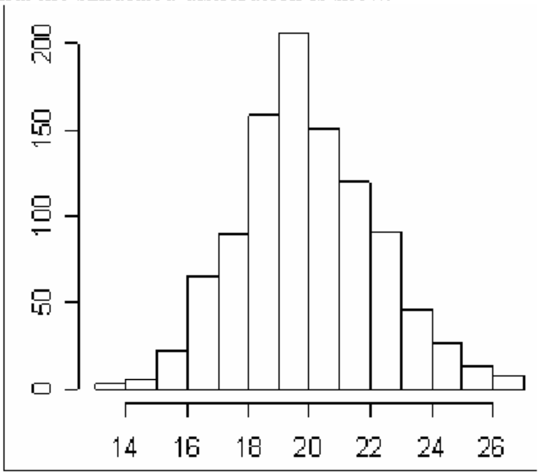



**Figure 4: Histogram of a sampled profit from one run**

(ii) Triangular distribution
In case of a triangle-probability function, ROI is symmetrically distributed in the interval [0.24, 0.26] and capital is also symmetrically distributed, however, in the interval [76, 84]. The proposal distribution is an uniform distribution. The other parameters do not differ from (i). Figure 5 shows the simulated profit. As before the simulated data differs clearly from the theoretical distribution at the boundaries. The expected mean of profit is $\mu = 20$ using equation (2) and the simulated value is $mean(\hat{\mu}) = 20$. The variance calculated by using equation (5) is $\sigma^2 = 0.275$ which corresponds to $\hat{\sigma}^2 = 0.215$. The theoretical value for the lower bound is $l = 18.24$ and for the upper bound $u = 21.84$. The minimum of the simulated data is $\min(\hat{l}) = 18.29$ and the maximum is $\max(\hat{u}) = 21.77$.

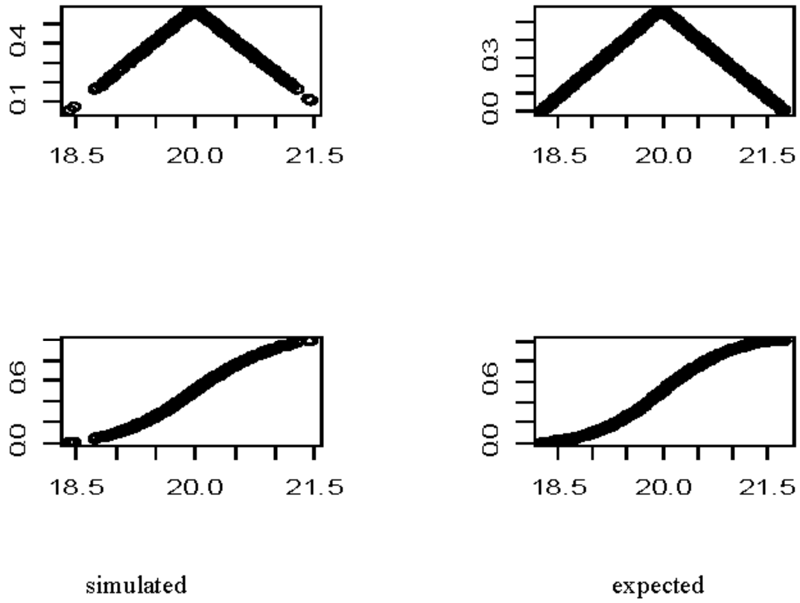



**Figure 5: Simulated and expected triangle distributions**

If the endogenous variables are described by a triangle-probability function, the simulated data for the exogenous variable should follow a triangle distribution. The two-sided Kolmogorov-Smirnov test has a mean in the p-values of 0. Thus the hypothesis, that the profit is described by a symmetrically triangle-distribution in the interval [18.24, 21.84] is rejected.

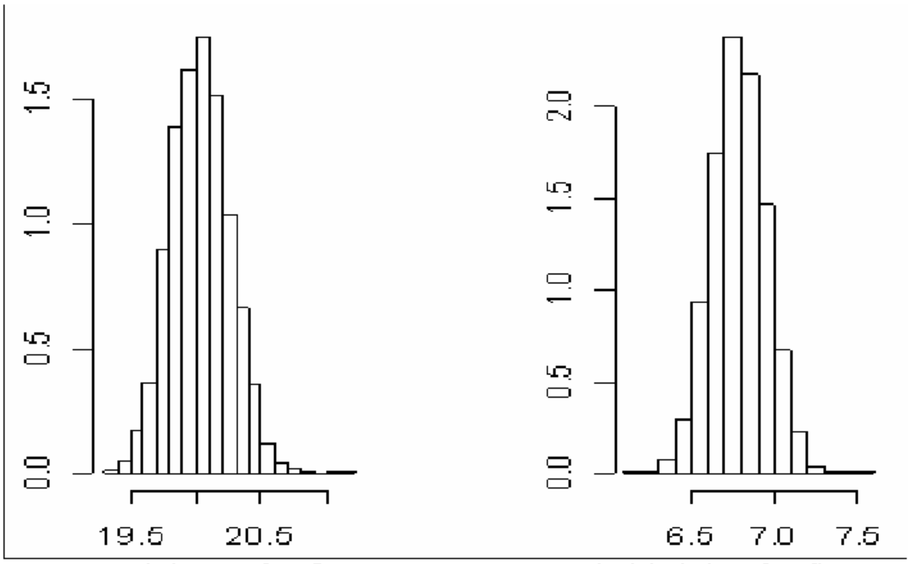
(iii) Contaminated Gaussian distribution
The next example the variables ROI and capital are distributed as a two-peak Gaussian which is described by equation (7):

ROI: $\mu_1 = 0.21 \quad {\sigma_1}^2 = 0.05^2 \quad \mu_2 = 0.31 \quad {\sigma_2}^2 = 0.06^2$ if $\varepsilon = 0.4$.

Capital: $\mu_1 = 71 \quad {\sigma_1}^2 = 10^2 \quad \mu_2 = 86 \quad {\sigma_2}^2 = 8^2$ if $\varepsilon = 0.6$.

The proposal distribution is a Gaussian distribution. From that distribution samples can be more easily drawn and thus the simulation needs less time. The simulation parameters remain as before.

The mean of the simulated data for profit is $avg(\hat{\mu}) = 20.03$. Equation (2) gives a mean of $\mu = 20$ and equation (5) gives a variance $\sigma^2 \approx 6.19$. The variance of the simulated profit is $\hat{\sigma}^2 = 6.78$. Figure 6 shows a distribution of the simulated mean and the simulated standard deviation of profit.




sampled mean of profit standard deviation of profit

**Figure 6: Histogram of the sampled mean and standard deviation of profit**

3.2.3. Folding two distributions under Division

The second non-linear operation is described in the following part. An equation with this operation is:

$$profit\ \ margin = profit / transaction\ \ volume.$$

Profit and transaction volume are the exogenous variables and the profit margin is the endogenous.

(i) Gaussian distribution

In the first case, profit and transaction volume are described by a Gaussian distribution function. The exogenous variables are distributed with

$$profit \sim N(20, 2^2) \quad \text{and} \quad transaction\ \ volume \sim N(100, 5^2).$$

The proposal distribution is selected also a Gaussian distribution in order to reduce the simulation size. The parameters of particles and repetition are the same as before.

The mean of the simulated data of profit margin is $\hat{\mu} = 0.20044$ that corresponds to the theoretical estimated value of equation (3) which is $\mu = 0.2005$. The standard deviation, which can be estimated using equation, (6) is $\sigma = 0.0005$, that is faced with a standard deviation of $\hat{\sigma} = 0.0005$ from the simulated data for profit margin.

Because of these results, a Kolmogorov-Smirnov goodness of fit test is implemented with the theoretical mean and variance for each run. In order to test the simulated data, however, a standardization of the above distribution to $N(0,1)$ is applied. Because division of two independent standard Gaussian distributed variables produces a Cauchy-distribution with parameters median equal to 0 and λ equal to 1, cf. Müller (1983). The mean of the computed p-values of the two-sided test is 0.548. So the $H_0$-Hypothesis cannot be rejected and the simulated data support a Cauchy distribution with the theoretical computed mean and variance. Figure 7 shows the simulated data of profit margin, profit and transaction volume selected from one run.

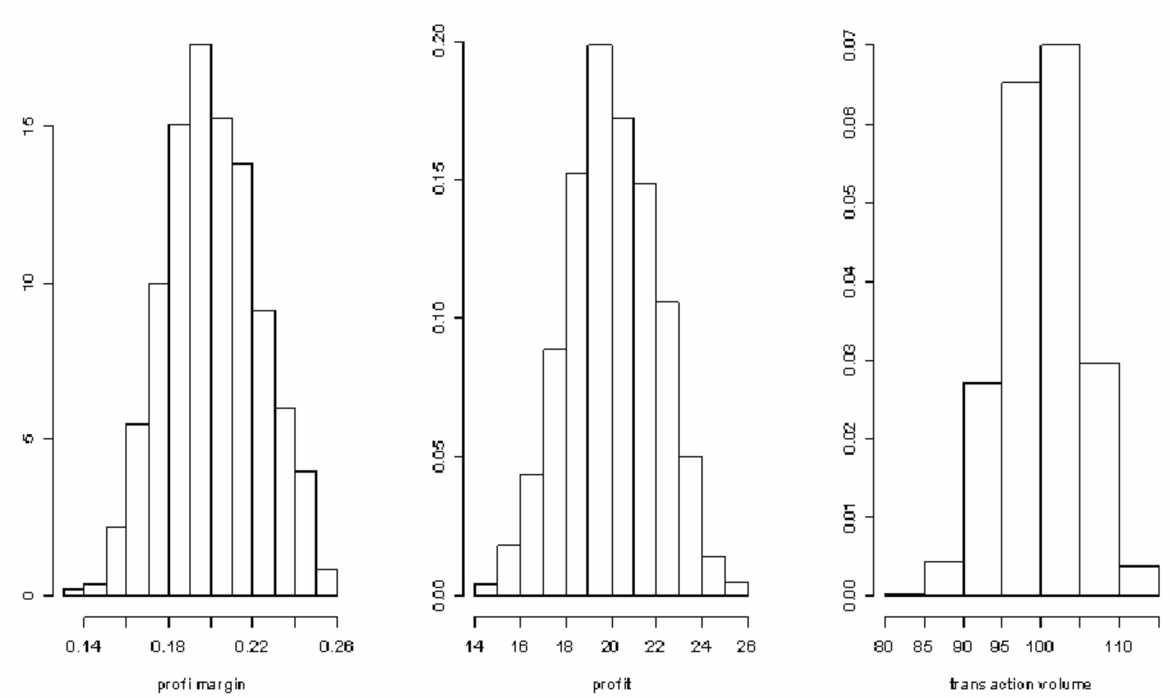



**Figure 7: Sampled distributions from all variables in profit margin = profit / transaction volume (Gaussian distributed variables)**

(ii) Triangle distribution
If both endogenous variables are symmetrically distributed with a triangle function the following results arise:
Profit is distributed in the interval [18, 22] and transaction volume in the interval [90, 110]. The proposal distribution is an uniform distribution for both variables. The algorithm is run 5000 times with 1000 particles. The simulated mean of profit margin is $avg(\hat{p}) = 0.2003$ and the standard deviation is $\hat{\sigma} = 0.102$. Estimating the mean by using equation (3) results in $\mu \approx 0.2003$ and the standard deviation can be approximated by equation (6) with $\sigma \approx 0.116$. A Kolmogorov-Smirnov test is in this case is impossible, because the 2nd moment does not exist.

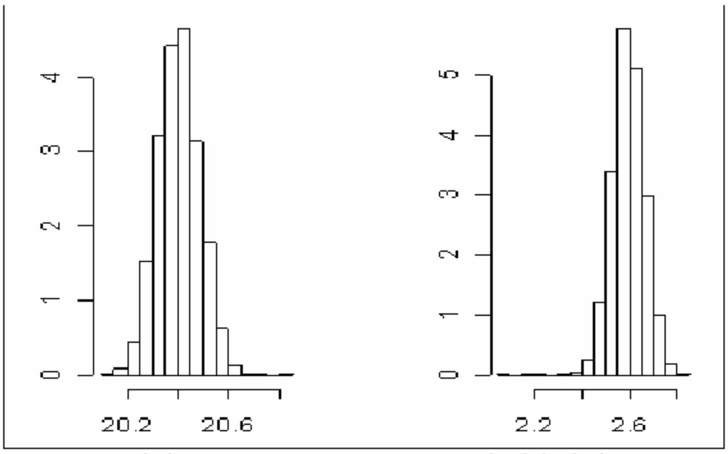
(iii) Contaminated Gaussian distribution
As a last example profit and transaction volume are distributed as described in equation (7).

Profit: $\mu_1 = 21 \quad \sigma_1^{\,2} = 3 \quad \mu_2 = 20 \quad \sigma_2^{\,2} = 2$ if $\varepsilon = 0.6$.

Transaction volume: $\mu_1 = 90 \quad \sigma_1^{\,2} = 3 \quad \mu_2 = 105 \quad \sigma_2^{\,2} = 5$ if $\varepsilon = 0.7$.

The proposal distribution for both variables is a Gaussian and all other simulation parameters are as above.

The simulated profit margin has a mean of $avg(\hat{\mu}) = 0.2037$ and a standard deviation of $\hat{\sigma} = 0.0312$. The estimation of the mean is $\mu \approx 0.204$ by using equation (3) and equation (6) calculates the estimation of the standard deviation with $\sigma \approx 0.0299$. Figure 8 shows the histogram of the simulated mean and of the simulated standard deviation for profit margin.




sampled mean standard deviation

**Figure 8: Distribution of mean and standard deviation of profit margin (contaminated Gaussian distribution)**

## 4. Variance-reduction Techniques

The problem of too large deviations caused by simulation can be reduced by robust estimation. To achieve robustness the extreme values at both ends must be eliminated from the sample. A robust method is the γ-trim mean and γ-trim variance. This implies to drop the g upper and g lower sampled values. The parameter g describes the amount of data which will be taken out from sample of size R. The estimator for the mean is changed to:

$$\hat{\mu}_\gamma = \frac{1}{R - 2 \cdot g} \sum_{i=g+1}^{R-g} X_i \quad \text{with } g = \gamma \cdot R \text{ and } 0 \le \gamma < 0.5 . \tag{8}$$

Accordingly, the estimator for the variance is:

$$\hat{\sigma}_\gamma^{\ 2} = \frac{1}{R - 2 \cdot g - 1} \sum_{i=g+1}^{R-g} \left( x_i - \hat{\mu}_\gamma \right) . \tag{9}$$

Because the sample size is reduced, sampling should be extended. The amount of post sampled particles depends on the target sampled size and the γ-trim. The relationship can be described by: $sampling\ \ particles = demanded\ \ particles / (1 - 2 \cdot \gamma)$. The γ parameter is dependent on the variance of the variable. If the variance is high, γ should be increased. On the other hand, the parameter should be set to 0 for a low variance. This will reduce the sampling effort and, consequently, speed-up the algorithm.

Our new MH algorithm with trimming has the following steps:

1. Initialization of the given exogenous variables with proposal distributions and desired probability functions. The parameter for trimming has to be set and the sampling size determined
2. Use the Metropolis-Hastings algorithm to sample from the distribution of the exogenous variables
3. Derive the distribution of the endogenous variables via the equation-system
4. Calculate the means and variances of all variables by using estimators from equation (8) and (9)
5. If the repetition size is not reached run by run, go to 2

## 5. Simulation of a system of equations

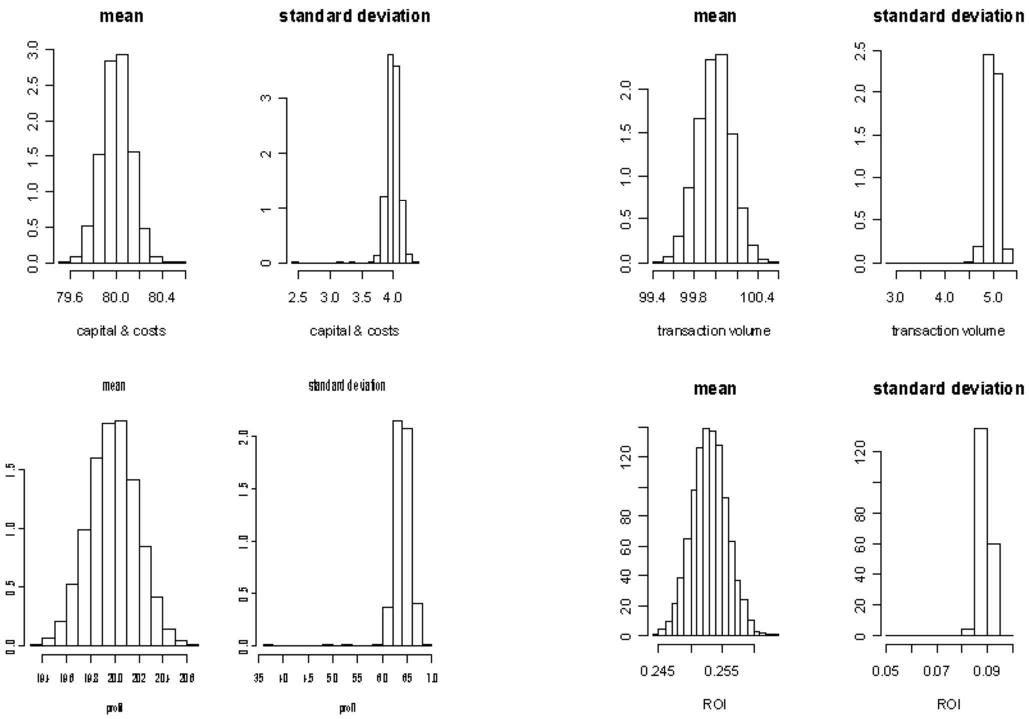
The following example illustrates, that not only a single equation but a full system of equations can be simulated by the algorithm “Metropolis-Hastings with Trimming”.
The exogenous variables are capital, costs and transaction volume and they are distributed as follows:

$$capital \sim N(80,4^2),\; costs \sim N(80,4^2) \text{ and } transaction\ \ volume \sim N(100,5^2)$$

To keep the simulation size small, the proposal distribution for the endogenous variables is also a Gaussian distribution. The parameters of simulation experiments are as before.
The following equations constitute the system of equations:

$$\begin{aligned} profit &= transaction\ \ volume - costs \\ return\ \ on\ \ investment &= profit\ /\ capital \\ profit\ \ margin &= profit\ /\ transaction\ \ volume \\ capital\ \ turnover &= transaction\ \ volume\ /\ capital \end{aligned}$$

Figure 9 shows the distributions of mean and standard deviation of all variables. The sampling is done with $\gamma$ equals to 0 (no trimming).



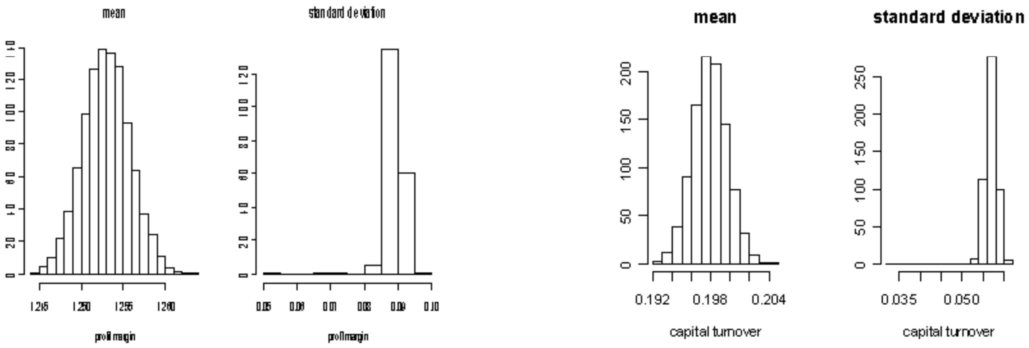
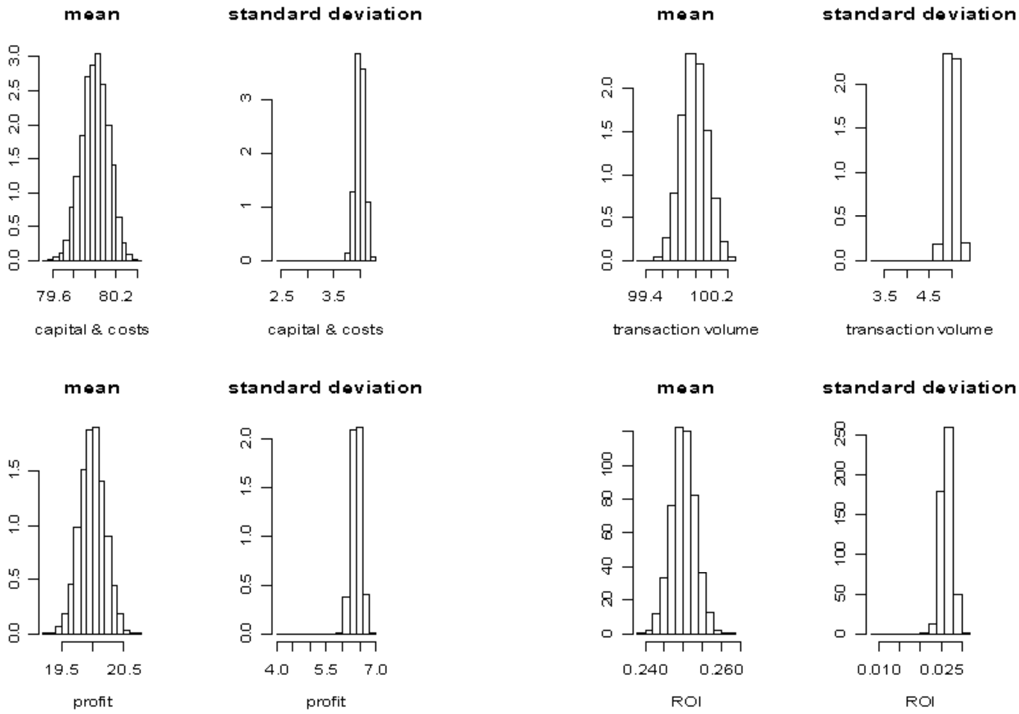



**Figure 9: Simulated means and standard deviations of the DuPont system (no trimming allowed)**

Figure 10 is the version where the $\gamma$-trimmed tough estimation is used. The value of $\gamma$ is set to 0.3. The shrinkage of the spread of all distributions becomes evident, cf. Figure 9 and 10.



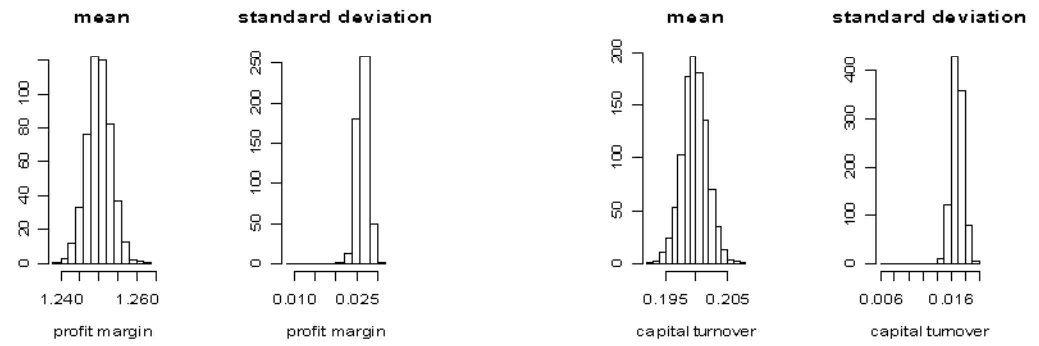



**Figure 10: Simulated means and standard deviations of $\gamma$-trimmed DuPont System ( $\gamma$ = 0.3 trimming set)**

## 6. Results and Perspectives

We showed that simulation based upon the Metropolis-Hastings algorithm is a sound method to evaluate multi equation systems. The necessary inputs are the system of equations and the probability functions of the exogenous variables. All other variables can be simulated. If the sampling size is adequate, the sampling distributions are similar to the theoretical distributions. Most of those theoretical distributions are not analytically derivable, thus sampling is the only way to solve such systems.
The DuPont system can be used for different controlling functions in a company. The improvement of using random variables instead of crisp numbers is obvious, if measurement errors of various kinds are effective. The controller can use simulation to accept or reject provided data given his domain knowledge as hypothesis.
The results of our simulation study of non-linear equation systems like the DuPont system are:

- Using the Metropolis-Hastings algorithm is a very flexible method to sample from any probability function. Moreover, these samples can be further used.
- A critical drawback is the sampling from a probability density function defined on a finite interval.
- If the variance is large, the MH algorithm should be modified to MH with trimming. The simulation size must be adapted. The pay-off for increased simulation efforts are improved estimators.

However further challenges still exist, and more research on simulation of closed intervals is needed. Another important area is the similarity between sampled distributions. Equally important issue is to investigate the sampling procedure when an equation system is over-determined.

## Acknowledgement

The third author thanks P. Cheeseman, NASA Ames Research Center, for drawing his attention to particle filters and Metropolis-Hastings procedures, and his perceptual propagation of the dominance of probability theory in science and AI.

## References

Büning, H. (1991). *Robuste und adaptive Tests*. Berlin: Walter de Gruyter.

Chib, S. (2004). Markov Chain Monte Carlo Technology. In J. E. Gentle, W. Härdle & Y. Mori (Ed.), Handbook of Computational Statistics, Concepts and Methods (pp. 71-102). Berlin, Heidelberg, New York: Springer Verlag.

Hastings, W.K. (1970). Monte Carlo sampling methods using Markov Chains and their applications. *Biometrica, 57 (1),* 97-109.

Mood, A. M., Graybill, F. A., & Boes, D. C. (1973). Introduction to the theory of Statistics (3rd ed.). Tokyo: McGraw-Hill.

Müller, P. H. (Ed.). (1983). Lexikon der Stochastik (4th ed.), Berlin: Akademie-Verlag.