

*The Impact of Clustering Method for Pricing a Large Portfolio of VA Policies*

By

*Zhenni Tan*

A research paper presented to the

University of Waterloo

In partial fulfillment of the requirements for the degree of

Master of Mathematics

in

*Actuarial Science*

Waterloo, Ontario, Canada, *2017*

©(Zhenni Tan) 2017

## **AUTHOR'S DECLARATION**

I hereby declare that I am the sole author of this thesis. This is a true copy of the thesis, including any required final revisions, as accepted by my examiners.

I understand that my thesis may be made electronically available to the public.

## **Abstract**

The purpose of this article is to find an efficient way to price a large portfolio of variable annuity contracts. We achieve this goal by improving a three-step procedure proposed by Gan (2013). The three-step procedure includes:(1)the  $k$ -prototypes method; (2)Monte Carlo Simulation for Variable Annuities; (3) The Kriging Method. The three-step method in Gan (2013) is studied and reproduced in this article. Additionally, three methods to identify representative contracts are considered, analyzed, and tested. The performances of these alternatives are measured by the speed and the accuracy.

## Table of Contents

AUTHOR'S DECLARATION.....	iii
Abstract.....	iii
List of Figures .....	iv
List of Tables .....	v
1 Variable Annuity .....	1
1.1 Guaranteed Minimum Benefits .....	2
1.1.1 Guaranteed Minimum Death Benefit.....	3
1.1.2 Guaranteed Minimum Withdrawal Benefit.....	3
2 The Three-step Method of Pricing a Large Portfolio of VA Policies .....	4
2.1 the $k$ -prototypes method.....	4
2.2 Monte Carlo Simulation for Variable Annuities .....	8
2.3 The Kriging Method .....	11
3 Proposed Alternatives to Step 1 .....	13
3.1 Nearest neighbor mapping across the subset(MAS) and mapping within the subset (MWS).....	14
3.2 Rounding of Discrete Attributes (RDA).....	15
3.3 Simple Random Sampling (SRS).....	16
4 Numerical Experiment Set.....	16
5 Test Results.....	17
6 Conclusion .....	26

Reference.....28

## List of Figures

Figure 1.....	8
Figure 2.....	15
Figure 3.....	19
Figure 4.....	21

## List of Tables

Table 1.....	18
Table 2.....	18
Table 3.....	22
Table 4.....	22
Table 5.....	23
Table 6.....	25



# The Impact of Clustering Method for Pricing a Large Portfolio of VA Policies

## 1 Variable Annuity

Variable annuities(VAs) are popular life insurance products in the insurance industry. VAs are long-term contracts designed for offering post-retirement income (Bacinelloa, Millosovich, Olivieri, & Pitacco, 2011). VAs have attractive features such as dynamic investment opportunity and guarantees against the investment and early death risks.

A single premium or recurrent premiums from policyholders are invested into the different risk return portfolio decided by the policyholders. There are different reference funds in each different risk return portfolio. Different risk return portfolio can be changed at no cost under some constraint conditions (for example, no more than once a year). Fees of additional guarantees, asset management, administration and expenses are charged by reduction of the account value (Bacinelloa, Millosovich, Olivieri, & Pitacco, 2011). Also, policyholders can surrender the contract, withdraw a portion of account value and annuitize the account value after some specific time. Many guarantees were first introduced in the 1990s (Bauer, Kling, & Russ, 2008)

The VAs combine funds with one or more guarantees. Without a guarantee, VAs are same as investment funds for policyholders and there is no risk for the insurer, comparing with the traditional life insurance. “Insurers act as a steward of the investment funds “(Hardy, 2003). Guarantees transfer the risk from the policyholders to the insurers, so the risk management and pricing of VAs is very important. For example, Equitable Life, one of the world oldest insurer, had sold a product with the option of a 7% guaranteed annuity rate at maturity in 1970s. In 1993, the annuity rate at that time became lower than the initial guaranteed rate and policyholders exercised that guarantee. The liability of Equitable Life increased massively

(McNeil, Frey, & Embre, 2015). In this case, the insurer underestimated the risk of this guarantee.

Except the financial risk, insurer also faces the longevity and policyholder behavior risks, which are same as the traditional life insurance. However, these risks interact each other and the interaction is not fully understood. Appropriate risk management is required. Pricing and hedging of guarantees are the major concern of the contract design. This risk should be considered during the process of pricing. For example, Bacinelloa, Millosovich, Olivieri and Pitacco(2011) introduced a unifying valuation approach for VAs depending on different assumptions of policyholder behavior

## **1.1 Guaranteed Minimum Benefits**

There are two main classes of guarantees: Guaranteed Minimum Death Benefit (GMDB) and Guaranteed Minimum Living Benefit (GMLB). There are three types of benefits in GMLB: Guaranteed Minimum Accumulation Benefit (GMAB), Guaranteed Minimum Income Benefit (GMIB), Guaranteed Minimum Withdrawal Benefit (GMWB).

The simplest form of guaranteed benefit is Guaranteed Minimum Accumulation Benefit, it provides a minimum account value at maturity. Usually, the minimum account value is the initial premium amount. It gives the policyholders the chance of renewing account value at maturity when the investment performance is bad.

Guaranteed Minimum Income Benefit also provides a guaranteed account value at maturity, but policyholders only can take out the account value in the form of an annuity at a specified

annuity rate. This guarantee also reduces the risk from the annuity rate at maturity. In the money means that the annuity payment resulting from converting the guaranteed account value at a specified annuity rate is greater than the actual account value at current annuity rate. Usually, policyholders can annuitize the real account value during a time period, and the predefined annuity rate for the guaranteed account value is always conservative, that the offered roll-up rates are greater than the risk-free rate of interest. Because of this, the option may not be in the money even if the guaranteed account value is greater than the actual account value. The policyholders will not exercise it (Bauer, Kling, & Russ, 2008).

### **1.1.1 Guaranteed Minimum Death Benefit**

With the Guaranteed Minimum Death Benefit, the beneficiary can receive a guaranteed amount if the policyholder dies before maturity. There is a variety of designs for Guaranteed Minimum Death Benefit: Return of Premium, Annual Roll-Up, Annual Ratchet, and Annual Reset (Bacinelloa, Millosovich, Olivieri, & Pitacco, 2011).

$G_t^D$  is the guaranteed amount,  $t$  is the death time before the maturity  $T$ . The guaranteed amount is the single premium paid.  $P$  is the premium.

$$G_t^D = P$$

This is the basic form of a death benefit, which is called Return of Premium Death Benefit. Return of Premium Death Benefit is used in this article. There are some other types of GMDB are not used in this article. However, some further research about pricing large portfolio may be designed for these more complex types of benefits, which are annual roll-up, annual ratchet and annual reset.

Policyholders receive the maximum of guaranteed amount and account value. The benefit is  $b_t^D = \max(A_t, G_t^D)$ ,  $t \leq T$ . If there is no guaranteed amount,  $G_t^D = 0$ , the policyholders will receive the account value

### **1.1.2 Guaranteed Minimum Withdrawal Benefit**

With a GMWB option, policyholders can withdraw money from the policy account even if account value is zero. The account value may decrease to zero because of bad investment performance or the longevity of the policyholders. The guaranteed withdrawal amount is the minimum of the remaining total amount that can be withdrawn and the maximum amount that can be withdrawn annually. The remaining total amount is  $G_t^W$ , which is a specified amount like single premium. The maximum annual withdrawal is a certain proportion  $x_w$  of  $G_0^W$ , which is  $G_t^E = x_w G_0^W$  (Bauer, Kling, & Russ, 2008).

There are some special forms of GMWB, which are not used in this article. However, some further research about pricing large portfolio may be designed on these more complex types of benefits, which are the step up GMWB and the GMWB for life.

## **2 The Three-step Method of Pricing a Large Portfolio of VA Policies**

The following section reviews the three-step method of pricing a large portfolio of VAs in Gan (2013). The Three-step are (1) obtaining  $k$  centroids by the  $k$ -prototypes method, (2) obtaining the values of centroids by the Monte Carlo simulation, and (3) obtaining the value of the large portfolio by the kriging method (Isaaks & Srivastava, 1990).

## 2.1 The k-Prototypes Method

Data clustering is a method used to group items into different clusters. The items in the same cluster are similar and the items in different clusters are dissimilar. Huang (1998) introduced a  $k$ -prototypes algorithm that allows for clustering objects with mixed numeric and categorical attributes. The  $k$ -prototype algorithm can be used to cluster a large portfolio of the VA contracts with mixed numeric and categorical attributes.

We define a large portfolio that contains  $n$  VA contracts as  $X = (\vec{x}_1, \vec{x}_2, \vec{x}_3, \dots, \vec{x}_n)$ . We denote the  $i$ th contract in the portfolio as  $\vec{x}_i$ . Attributes of each contract are classified into numerical attributes and categorical attributes. Assuming there are  $d$  attributes, the  $d_1$  attributes are numerical, the  $d_2 = d - d_1$  attributes are categorical. Then the distance between two contracts' attributes is,

$$D(\vec{x}, \vec{y}, \lambda) = \sqrt{\sum_{h=1}^{d_1} (x_h - y_h)^2 + \sum_{h=d_1+1}^d \delta(x_h, y_h)}$$

$$\delta(x_h, y_h) = \begin{cases} 0, & \text{if } x_h = y_h \\ 1, & \text{if } x_h \neq y_h \end{cases}$$

The  $x_h$  is the  $h$ th attribute of contract  $\vec{x}$ . The  $\lambda$  is a parameter used to balance the weight of numerical attributes and categorical attributes.

All numerical attributes in the distance function should be normalized. The numerical attributes minus the mean and divide the standard deviation of  $x_{1h}, x_{2h}, x_{3h}, \dots, x_{1nh}, h = 1, 2, 3, \dots, d_1$ . Otherwise, the distance will be almost only influenced by the attributes from premium, which is a very large number comparing with other attributes.

However, there are some shortcomings for this distance function. Firstly,  $\delta(p, q) = 0$  for  $p = q$  and  $\delta(p, q) = 1$  for  $p \neq q$  are used for categorical attributes, it does not reflect the real situation appropriately. Stanfill and Waltz(1986) observed that it is true that  $\delta(p, q) = 0$  for  $p = q$ , but it is not necessarily true that  $\delta(p, q) = 1$  for  $p \neq q$  in supervised learning. Secondly, the  $\lambda$  is user-defined, and all numerical attributes have equal weight. An inappropriate  $\lambda$  may lead to inaccurate clustering and numerical attributes may not have equal effect on clustering (Ahmad & Dey, 2007).

The  $k$ -prototype algorithm tries to minimize the function  $P_\lambda = \sum_{j=1}^k \sum_{\vec{x} \in C_j} D(\vec{x}, \vec{\mu}_j, \lambda)$ , where  $D(\cdot)$  is the distance function,  $k$  is the number of centroids and  $C_j$  is the  $j$ th cluster,  $\vec{\mu}_j$  is the centroids of the  $C_j$  cluster. We call the contracts that belong to a cluster a member of that particular cluster. The  $k$ -prototype algorithm can update the cluster memberships  $\gamma_i^{(0)}$  given the centroids, and it can update the centroids given cluster memberships. This process is repeated until some conditions (explained in step1.4) happened. This process includes the following steps,

**Step 1.1:** Initialize the centroids

$K$  distinct contracts are randomly chosen from the portfolio as an initial centroid, which is

$$\vec{\mu}_1^{(0)}, \vec{\mu}_2^{(0)}, \vec{\mu}_3^{(0)}, \dots, \vec{\mu}_k^{(0)}.$$

**Step 1.2:** Update the clustering memberships

Then, the clustering memberships are updated  $\gamma_1, \gamma_2, \gamma_3, \dots, \gamma_n$  as,

$$\gamma_i^{(0)} = \operatorname{argmax}_{1 \leq j \leq k} D(\vec{x}, \vec{\mu}_j^{(0)}, \lambda),$$

Where  $D(\cdot)$  is the distance function.

**Step 1.3:** Update the Centroids.

The centroids are updated at this step as,

$$\tilde{\mu}_{jh}^{(1)} = \frac{1}{C_j} \sum_{\tilde{x} \in C_j} x_h, h = 1, 2, 3, \dots, d_1$$

$$\tilde{\mu}_{jh}^{(1)} = \text{mode}_h(C_j), h = d_1 + 1, \dots, d$$

where  $C_j = \{\tilde{x}_i \in C_j : \gamma_i^{(0)} = j\}$  for  $j = 1, 2, 3, \dots, k$  and the  $\text{mode}_h(C_j)$  is the value of the  $h$ th categorical attributes, which appeared most frequently. The  $h$ th attribute can take  $m_h$  different values  $A_{h1}, A_{h2}, A_{h3}, \dots, A_{hm_h}$ . The  $f_{ht}(C_j)$  is how many times the  $h$ th attribute equal to values  $A_{ht}, t = 1, 2, 3, \dots, m_h$ , which is,

$$f_{ht}(C_j) = |\tilde{x} \in C_j : x_h = A_{ht}|, t = 1, 2, 3, \dots, m_h.$$

Then,

$$\text{mode}_h(C_j) = \underset{1 \leq t \leq m_h}{\text{argmax}} f_{ht}(C_j), h = d_1 + 1, \dots, d$$

**Step 1.4:** Repeat Step2 and Step 3 until the stop condition is satisfied.

The stop condition is there is no change for the centroids or memberships after the update, or it reaches the maximum iteration times.

**Step 1.5:** Nearest neighbor mapping

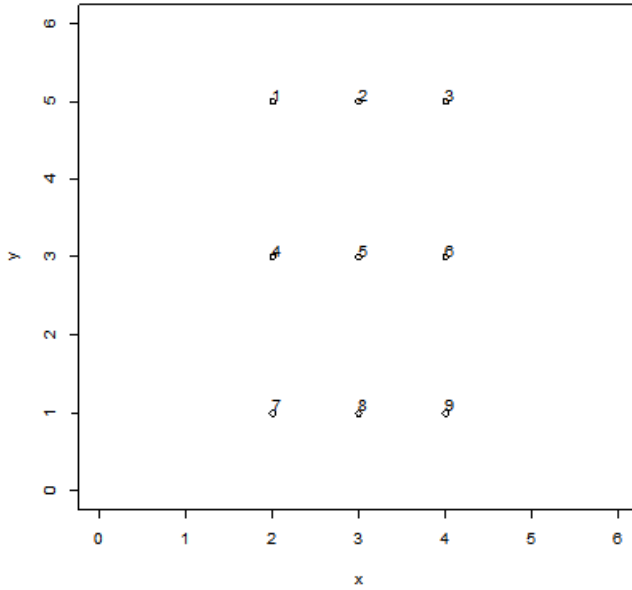
$k$  centroids are obtained by  $k$ -prototype method, and the centroids are denoted as  $\vec{\mu}_1, \vec{\mu}_2, \vec{\mu}_3, \dots, \vec{\mu}_k$ . Given the centroids, the representative VA contracts,  $\vec{z}_1, \vec{z}_2, \vec{z}_3, \dots, \vec{z}_k$ , can be selected as follows,

$$\vec{z}_j = \operatorname{argmax}_{\vec{x} \in X} D(\vec{x}, \vec{\mu}_j, \lambda)$$

The representative variable annuity contract  $\vec{z}_j$  is closest to the centroid  $\vec{\mu}_j$  in the total contract portfolio  $X = (\vec{x}_1, \vec{x}_2, \vec{x}_3, \dots, \vec{x}_n)$ . After this step, note that the duplication is checked to make sure that the VA representative contracts are different with each other, which is,  $D(z_r, z_s, \lambda) > 0$ , for all  $1 \leq r < s \leq k$ .

This  $k$ -prototypes algorithm could be slow because there are a lot of distance calculations with large  $k$  and  $n$ . In such case, according to the Gan (2013), the portfolio of VA contracts is first divided into subsets and then repeat the cluster process inside each subset. According to Gan (2013): “200000 contracts portfolio is divided into 33 subsets with 3 or 4 centroids in each subset.” Then 100 representative contracts are produced quickly. According to the Gan (2013): “if the portfolio is large and the policies are evenly distributed, it produces the similar result as the  $k$ -prototype algorithm without subsets”. However, it will produce the inaccurate result if the portfolio is small. According to an example in Gan (2013),  $\{1,2,3,4,5,6,7,8,9\}$  in Figure 1 is clustered into  $\{1,2,3\}, \{4,5,6\}, \{7,8,9\}$ . However, after we divide  $\{1,2,3,4,5,6,7,8,9\}$  into two subsets:  $\{1,2,3,4,5\}$  and  $\{6,7,8,9\}$ , then cluster the  $\{1,2,3,4,5\}$  into two centroids, and cluster the  $\{6,7,8,9\}$  into one centroid. Then, the result is  $\{1,2,3\}, \{4,5\}, \{6,7,8,9\}$ .





**Figure. 1**

## 2.2 Monte Carlo Simulation for Variable Annuities

After the representative contracts are selected after the clustering process, the value of the contracts can be evaluated by the Monte Carlo simulation for VAs in Bauer, Kling, Russ (2008). The following mathematical notations are used to describe the Monte Carlo method for pricing the GMDB rider and the GMWB rider.

The underlying mutual fund at time  $t$  of the VA is denoted by  $S_t$ . The account value at time  $t$  is denoted as  $A_t$ . The withdrawal benefit at time  $t$  is denoted as  $W_t$ . The death benefit at time  $t$  is denoted as  $D_t$ . The remaining total amount can be withdrawn and the maximum amount can be withdrawn annually are denoted as  $G_t^W$  and  $G_t^E$ . The guaranteed death benefit is denoted as  $G_t^D$ . The certain proportion of the premium  $G_0^W$  is denoted as  $x_w$ , that can be withdrawn annually  $G_t^E = x_w G_0^W$ . The maturity time of the contracts is  $T$ .

In this article, the impact of clustering in the three-step method is discussed in test Section 3. Since Gan (2013) only considered the GMDB rider and the GMWB rider, the same assumptions are applied in this article. There are two possible events of policyholders under this assumption; the policyholders withdraw money or dies.

The value immediately before and after the events happened at time  $t$  are denoted as  $(\cdot)_t^-$  and  $(\cdot)_t^+$ .

In this article, the Black-Scholes model is used to model the underlying mutual fund  $S_t$ . For the Black-Scholes model, the underlying mutual fund of the variable annuity is simulated as,

$$S_0 = 1, S_t = S_{t-1} \exp\left(\left[r - \frac{\sigma^2}{2}\right] + \sigma Z\right)$$

Where the  $r = 3\%$  denotes the interest rate and the  $\sigma = 20\%$  denotes the volatility of the underlying mutual fund  $S_t$ ,  $Z$  denotes a standard normal random variable. 1000 paths of  $S_t$  are simulated.

The following policyholder behavior is assumed: the policyholders take the maximum available withdrawals annually; All events occur at anniversary date  $t$  without fees and lapses; The 1996 IAM mortality Tables for males and females from SOA are used in the Monte Carlo process.

The initial values of guaranteed benefits are,

$$G_0^W = A_0, G_0^E = x_w A_0, \text{ and } G_0^D = A_0$$

The development of account value between  $(t)^+$  and  $(t + 1)^-$  is,

$$A_{t+1}^- = A_t^+ \frac{S_{t+1}}{S_t}$$

During the development between  $(t)^+$  and  $(t + 1)^-$ , the guaranteed minimum benefit does not change during the year if the corresponding guaranteed benefits are not roll-up base, so  $G_{t+1}^{W-} = G_t^{W+}$ ,  $G_{t+1}^{E-} = G_t^{E+}$ ,  $G_{t+1}^{D-} = G_t^{D+}$ .

At the anniversary date  $(t + 1)$ , the additional death benefit with a GMDB, comparing with the contracts without a GMDB, is,

$$D_{t+1} = \max(0, G_{t+1}^{D-} - A_{t+1}^-).$$

If the account value is greater, the policyholders receive the account value and the additional benefit is zero. If the guaranteed death benefit of the contract with a GMDB is greater than the account value, the policyholders exercise the option and gets the additional benefit  $G_{t+1}^{D-} - A_{t+1}^-$ .

The guaranteed withdrawal amount is the minimum of remaining total amount and annually withdrawal amount. The guaranteed withdrawal amount is,

$$E = \min(G_{t+1}^{W-}, G_{t+1}^{E-})$$

We assume the policyholders take the maximum available amount, so the additional withdrawal benefit with GMWB is,

$$W_{t+1} = \max(0, E - A_{t+1}^-)$$

If the account value is greater, the policyholders receive the account value and the additional benefit is zero. If the guaranteed withdrawal benefit of the contract with a GMWB rider is greater than the account value, the policyholders exercise the option and gets the additional benefit  $E - A_{t+1}^-$ .

If there is no GMWB rider, let  $x_w = 0$ . Then  $E = 0$ , and the additional withdrawal benefits  $W_{t+1} = \max(0, E - A_{t+1}^-) = 0$  without a GMWB.

The account value after  $t+1$  becomes  $A_{t+1}^+ = \max(0, A_{t+1}^- - E)$  after the events of withdrawals. The remaining total amount that can be withdrawn after  $t+1$  becomes  $G_{t+1}^{W+} = \max(0, G_{t+1}^{W-} - E)$ , and the maximum amount can be withdrawal annually,  $G_{t+1}^{E-}$ , does not change. The guaranteed minimum death benefit will be a pro rata scheme adjustment,  $G_{t+1}^{D+} = \frac{A_{t+1}^+}{A_{t+1}^-} G_{t+1}^{D-}$ . The pro rata scheme is used to avoid the adverse selection effects, and the withdrawals reduce the guaranteed death benefits. If  $A_{t+1}^- = 0$ , the Guaranteed death benefit becomes zero.

Then the value of a GMDB rider and a GMWB rider can be priced as the present value of the additional value of death and withdrawal benefits, comparing with the contracts without a GMDB rider and a GMDB. It is the value of the withdraw and death benefit in total.

$$V(S_1, S_2, S_3, \dots, S_{T+1}) = \sum_{t=1}^{T+1} {}_{t-1}p_{x_0} (1 - q_{x_0+t-1}) W_t e^{-rt} + \sum_{t=1}^{T+1} {}_{t-1}p_{x_0} q_{x_0+t-1} D_t e^{-rt}$$

The value of a GMDB rider and a GMWB rider by Monte Carlo method is the average of  $V(S_1, S_2, S_3, \dots, S_{T+1})$  along all paths.

## 2.3 The Kriging Method

After the value of representative contracts are calculated by the Monte Carlo method, the kriging method (Isaaks & Srivastava, 1990) is applied in Gan (2013) .

The representative VA contracts,  $\vec{z}_1, \vec{z}_2, \vec{z}_3, \dots, \vec{z}_k$  are selected by the clustering algorithm. Let  $y_j, j = 1, 2, 3, \dots, k$  be the fair value of  $\vec{z}_j$  which is priced by the Monte Carlo method. The kriging weights  $w_{i1}, w_{i2}, w_{i3}, \dots, w_{ik}$  can be used to estimate the fair value of VA contracts  $\vec{x}_i$  by the following formula,

$$\hat{y}_i = \sum_{j=1}^k w_{ij} \cdot y_j.$$

The kriging weights  $w_{i1}, w_{i2}, w_{i3}, \dots, w_{ik}$  can be calculated by the following linear equations,

$$\begin{pmatrix} V_{11} & \dots & V_{1k} & 1 \\ \vdots & \ddots & \vdots & \vdots \\ V_{k1} & \dots & V_{kk} & 1 \\ 1 & \dots & 1 & 0 \end{pmatrix} \begin{pmatrix} w_{i1} \\ \vdots \\ w_{ik} \\ \theta_i \end{pmatrix} = \begin{pmatrix} D_{i1} \\ \vdots \\ D_{ik} \\ 1 \end{pmatrix}$$

Where the  $\theta_i$  is a control variable, which is used to make sure that  $\sum_{j=1}^k w_{ij} = 1$ .

$$V_{rs} = \alpha + \exp\left(-\frac{3}{\beta} D(\vec{z}_r, \vec{z}_s, \lambda)\right), r, s = 1, 2, 3, \dots, k,$$

$$D_{ij} = \alpha + \exp\left(-\frac{3}{\beta}D(\vec{x}_i, \vec{z}_j, \lambda)\right), j = 1, 2, 3, \dots, k,$$

The  $D(\cdot)$  function is the distance function mentioned in the clustering section,  $\alpha \geq 0$  and  $\beta \geq 0$  are two parameters. The above linear equation system has a unique solution because  $D(\vec{z}_r, \vec{z}_s, \lambda) > 0$  for all  $r, s = 1, 2, 3, \dots, k,$ .

Then the fair value of whole portfolio  $X = (\vec{x}_1, \vec{x}_2, \vec{x}_3, \dots, \vec{x}_n)$  can be estimated by the following formula,

$$\hat{Y} = \sum_{i=1}^n \hat{y}_i = \sum_{i=1}^n \sum_{j=1}^k w_{ij} \cdot y_j = \sum_{j=1}^k w_j \cdot y_j$$

with

$$w_j = \sum_{i=1}^n w_{ij}$$

Then the efficient way of estimate the fair value of whole portfolio  $X(\vec{x}_1, \vec{x}_2, \vec{x}_3, \dots, \vec{x}_n)$  is solving the following linear equation system,

$$\begin{pmatrix} V_{11} & \dots & V_{1k} & 1 \\ \vdots & \ddots & \vdots & \vdots \\ V_{k1} & \dots & V_{kk} & 1 \\ 1 & \dots & 1 & 0 \end{pmatrix} \begin{pmatrix} w_1 \\ \vdots \\ w_k \\ \theta \end{pmatrix} = \begin{pmatrix} D_1 \\ \vdots \\ D_k \\ n \end{pmatrix}$$

Where the  $\theta$  is a control variable, which is used to make sure that  $\sum_{j=1}^k w_j = n$ , and  $D_j = \sum_{i=1}^n D_{ij}$ ,  $j = 1, 2, 3, \dots, k$ . Also, this linear equation system is the sum of both sides of the previous linear equation system from  $i = 1, 2, 3, \dots, n$ . It is more efficient because we only need

to solve one linear equation system instead of solving  $n$  linear equation systems and then sum the results.

### **3 Proposed Alternatives to Step 1**

In this section, some alternatives to step1 are introduced and tested. Clustering is a data mining technique that is used for observing interesting objects, and it is an unsupervised learning algorithm comparing with classification (Ahmad & Dey, 2007). More similar objectives are partitioned into same cluster, and interesting objectives may be discovered. The  $k$ -prototypes algorithm generalizes the  $k$ -means and  $k$ -modes algorithms, which is efficient in cluster large data sets with mixed numerical and categorical values.

The clustering part is time consuming in the three-step method for valuation of large portfolio of VAs at most of the time. Speed and accuracy are two important performance measures of the pricing method for a large portfolio. Different data clustering methods are considered.

Firstly, four different kinds of methods are tested in the first step: nearest neighbor mapping across the subset(MAS), mapping within the subset (MWS), rounding of discrete attributes (RDA) and simple random sampling (SRS).

#### **3.1 Nearest neighbor mapping across the subset(MAS) and mapping within the subset (MWS)**

The  $k$ -prototypes are used in the MAS, MWS and RDA method. The only difference is how the representative contracts are selected. The  $k$ -centroids  $\vec{\mu}_1, \vec{\mu}_2, \vec{\mu}_3, \dots, \vec{\mu}_k$  are obtained by  $k$ -prototype algorithm. Then, for the MAS, the representative contracts are selected from the whole portfolio. For the MWS method, the representative contracts are selected from each

subset of the portfolio  $X = (\vec{x}_1, \vec{x}_2, \vec{x}_3, \dots, \vec{x}_n)$ . For example, 100000 contracts are divided into 20 subsets, each subset has 5000 contracts. For the MAS method, we obtain 5 centroids in the subset, then choose 5 closest contracts from the whole portfolio  $X = (\vec{x}_1, \vec{x}_2, \vec{x}_3, \dots, \vec{x}_n)$  that contains 100000 contracts. For the MWS method, we obtain 5 centroids in the subset, then choose 5 closest contracts from the 5000 contracts in the same subset as representative contracts. The MWS method can save a lot of time comparing with the MAS method of avoiding a lot of distance calculations in the nearest neighbor mapping closest contract step. However, the closest contracts in the MAS method usually have shorter distances to the centroids. The 5000 contracts in each subset are also included in the whole portfolio  $X = (\vec{x}_1, \vec{x}_2, \vec{x}_3, \dots, \vec{x}_n)$ , and it is possible that exists contracts with shorter distances to the centroids from other subsets. This impact on accuracy is tested in Table 4 and Table 5. For example, in Figure 2, if the centroid is denoted as C and the closest contract in the same subset is B. The contract with shortest distance to the centroid is denoted as A, but it is not divided into same subset. As a result, the B is the representative contract by the MWS method, and the A is the representative contract by the MAS method. The closer total distance in MAS increases the accuracy of the result. However, the duplication produced in MAS decrease the accuracy. It is necessary to analyze which one has the stronger influence on the accuracy.



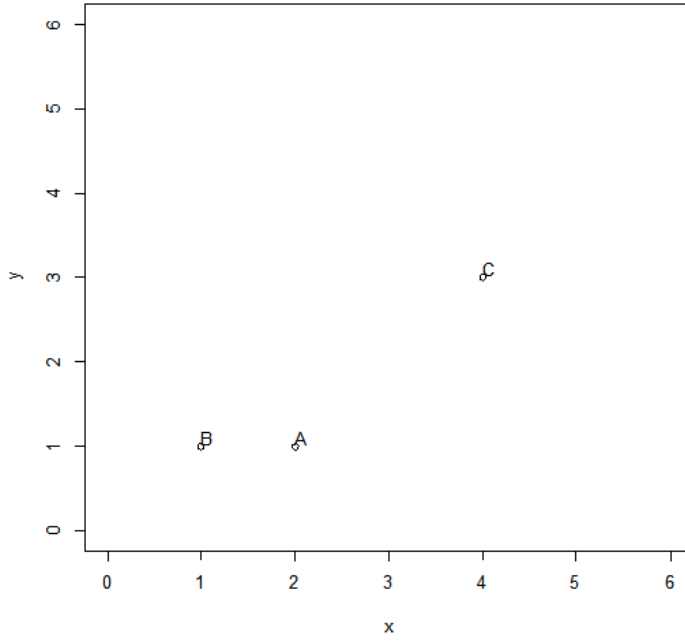


Figure. 2

### 3.2 Rounding of Discrete Attributes (RDA)

For the RDA method, the step of selecting closest contracts as representative contracts is removed, which can save a lot of time. The centroids are applied in the Monte Carlo step and kriging step directly. The numerical attributes of centroids  $\vec{\mu}_1, \vec{\mu}_2, \vec{\mu}_3, \dots, \vec{\mu}_k$  are normalized, which is the mean of the normalized numerical attributes from the contracts in the same cluster. However, the original numerical attributes are needed for Monte Carlo simulation to calculate the value of centroids. To estimate the original numerical attributes of centers, the mean and variance of numerical attributes are estimated from the whole portfolio  $X = (\vec{x}_1, \vec{x}_2, \vec{x}_3, \dots, \vec{x}_n)$ . For example, we can estimate mean  $\mu_{age}$  and standard deviation  $\sigma_{age}$  of age from the whole portfolio  $X = (\vec{x}_1, \vec{x}_2, \vec{x}_3, \dots, \vec{x}_n)$ . The age attribute in the centroid  $\vec{\mu}_1$  is denoted as  $\mu_{11}$ . The original age is  $\mu_{11} \times \sigma_{age} + \mu_{age}$ . Additionally, the age and maturity need to be rounded to the nearest integer after we get the original number. The results of the RDA method can be used

to compare with the results in the MAS method and the MWS method to see the influence of nearest contract mapping.

### **3.3 Simple Random Sampling (SRS)**

For the SRS method,  $k$  contracts are selected randomly from the whole portfolio without clustering at all. The resulting random sample with size  $k$  is used in the Monte Carlo step and kriging step. The results of the SRS method can be used to compare with the results in MAS, MWS, and RDA methods to study the influence of clustering in the pricing process. However, the influence of clustering the data in reality may be stronger than the results from the 100000 synthetic contracts created by uniform distributions.

Secondly, the different ways of dividing the whole portfolio of VA contracts into subsets are tested. The number of subsets are additional managerial decisions with different portfolio size, an appropriate way of dividing the whole portfolio can save time and improve accuracy of the results.

The result of a full-scaled Monte Carlo method with 100000 contracts and 10000 replications is used to measure the accuracy of the results by the tested methods. In each replication, one sample path of the underlying fund is simulated and the value of the VA contract is calculated accordingly. The number of paths increase to 10000, so the result of full-scaled Monte Carlo method is close to the true value of the large portfolio. The full-scaled MC result is denoted by subscript “MC”. Also, each tested method is repeated for 100 times to get the mean and variance of each result.

## 4 Numerical Experiment Set

The assumptions in the following table are used to create a large portfolio of synthetic VA contracts. 100000 contracts are created in this article.  $N$  is the natural number and  $R$  is the real number. Each contract  $\vec{x}_i = (x_{i1}, x_{i2}, x_{i3}, x_{i4}, x_{i5}, x_{i6})$  has six attributes: Maturity, Age, Premium,  $x_w$ , Gender and Type, hence the  $d_1$  is equal to 4, and  $d_2$  is equal to  $d - d_1 = 6 - 4 = 2$ .

Attribute	Values
<b>Guarantee type</b>	GMDB only, GMDB + GMWB
<b>Gender</b>	Male, Female
<b>Age</b>	$N \cap [20,60]$
<b>Premium</b>	$R \cap [10000,500000]$
<b>GMWB withdrawal rate</b>	0.04, 0.05, 0.06, 0.07, 0.08
<b>Maturity</b>	$N \cap [10,25]$

However, some assumptions about clustering part are changed to test the impact of clustering method for pricing a large portfolio of VA policies from Gan (2013).

## 5 Test Results

In Gan (2013), there is no description about how the nearest neighbor mapping is done, whether within subsets or across subsets. However, Gan (2015) discussed the “duplication”. Based on this, to our best understanding, the MAS method is used because it is the only method that will produce duplicated representative contracts. This speculation is explained after Table 3, because of this, the results of the MAS method are used as the MAS to which the other methods are compared.

The total value of the 100000 synthetic VA contracts by the MAS are produced for 100 times. The percentage error of the MAS estimator compared to the full-scaled MC estimator is calculated by  $\left| \frac{value_{MAS}}{value_{MC}} - 1 \right| \times 100$ ,

where  $value_{MAS}$  is the value by the MAS method, and  $value_{MC}$  is the value by the full-scaled Monte Carlo method with 10000 paths, which is close to the actual value of the large portfolio. The MAS method is repeated 100 times, then  $value_{MAS}$  of the 100 macro-replications are obtained. Also, percentage error of the 100 macro-replications are obtained. The mean and standard error of the percentage error can be calculated by the 100 macro-replications.

MAS % Error	100	500	1000	2000
1	43.15 (0.52)	13.92 (0.43)	8.61 (0.26)	5.37 (0.22)
2	17.74 (0.89)	5.56 (0.37)	2.87 (0.28)	2.29 (0.15)
5	9.94 (0.47)	3.96 (0.22)	2.02 (0.15)	0.99 (0.08)
10	3.88 (0.27)	4.27 (0.14)	3.3 (0.12)	1.73 (0.08)
50	1.93 (0.13)	2.4 (0.11)	2.62 (0.09)	1.88 (0.07)

**Table.1**

The percentage error by the MAS method are calculated. The column name is the total number of clusters. The row name is the number of cluster per subset. The number in the parentheses is the standard error of the 100 data.

MAS Total Time	100	500	1000	2000
1	4.56	18.46	36.05	73.35
2	7.17	22.99	43.34	84.60
5	22.84	34.99	53.69	93.26
10	40.40	44.19	62.07	102.041
50	176.42	105.57	113.06	143.29

**Table.2**

The total time (seconds) of the MAS method are presented .The column name is the total number of centroids. The row name is the number of cluster per subset. The number in the parentheses is the standard error .

From Table 1, the number of centroids per subset and the total number of centroids both have a strong impact on accuracy. The accuracy improves as the total number of centroids increases, which is similar to one of the finding in Gan (2013). Our experiment also reveals that increasing the number of centroids per subset can increase the accuracy of the result even when the total number of centroids is fixed.

In Table 2, the 100\*1 is the (total number of centroids)\* (the number of centroids per subset). The 100\*1 is the fastest. By increasing 1 centroid per subset to 10 centroids to subset, total time increase to the 40.40 seconds with the 3.88% error. By increasing 100 total number of centroids to 2000, total time increase to the 73.35 seconds with the 5.37% error, which spends more time with less accuracy. According to this, to increase the accuracy, increase the number of centroids per subset is more efficient than increase the total number of centroids, even both of these can increase the accuracy. More efficient means producing more accurate results with similar running times.

Increasing the number of centroids per subset can increase the accuracy of the result even when the total number of centroids is fixed. Unfortunately, we pay the price for this higher accuracy because increasing the number of centroids per subset also increase the time needed in the

clustering step. For example, when the total number of centroids is 100 and the number of centroids per subset is 25, there are less distance calculations comparing with the number of centroids per subset is 50, like Figure 3. The black space in Figure 3 represents the number of distance calculations. The area of black space in the right plot with 2 subsets is greater than the left plot with 4 subsets, which means it has more distance calculations. Additionally, increase the number of centroids per subset does not increase the time of Monte Carlo step and kriging weight step because the total number of centroids is fixed.

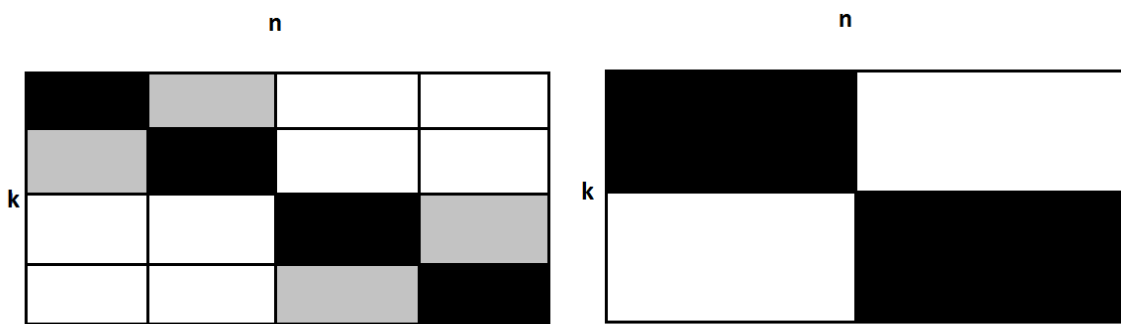
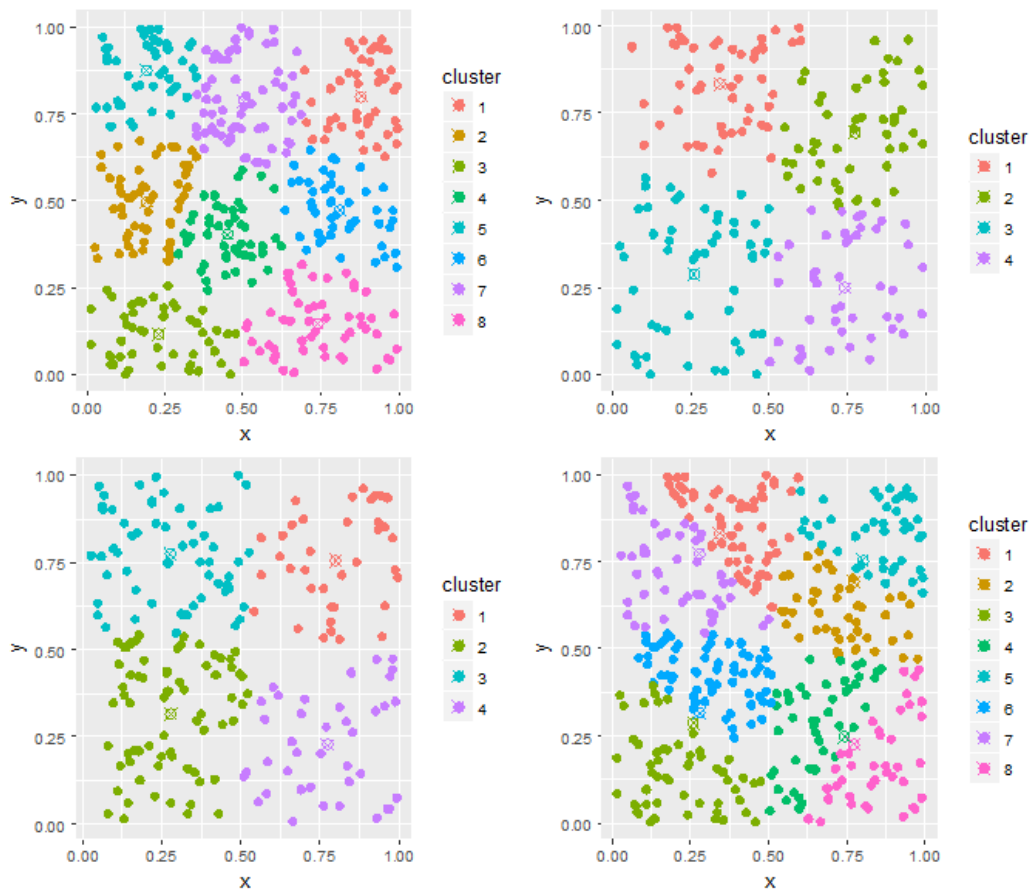


Fig.3

When the number of subsets increase, the decrease of the accuracy can be explained by Figure 4. Figure 4 is a simplified example of 400 points with two attributes, and there are 8 centroids in total. The contracts in this article are more complex, but it has the similar process of data clustering. The first plot is clustering without subset. Then the portfolio is divided into two subsets, the second plot is the first subset with half points selected from the whole portfolio randomly, and the third plot is the second subset. The fourth plot is the combination of the second plot and the third plot, it re-divides the whole portfolio into different clusters with the centroids produced by the second plot and third plot. After one portfolio is divided into two subsets, the total distance between the contracts to their centroids, which is in the fourth plot, is almost the double of the original total distance, which is in the first plot. The points in the cluster have longer distance to the centroids in the second or the third plot with 4 centroids than the first plot with 8 centroids. The points are uniformly distributed around the space, so

the total distance of first plot is same as the second plot, which is similar for the third plot. As a result, the total distance of the fourth plot, which is 14.00736, is close to the sum of the total distance in the second and the third plot, which is almost the double of the total distance in the first plot, which is 7.594413. The  $k$ -prototype algorithm tries to minimize the  $P_\lambda = \sum_{j=1}^k \sum_{\vec{x} \in C_j} D(\vec{x}, \vec{\mu}_j, \lambda)$ , and dividing subsets increase the total distance, which influence the performance of the clustering. The centroids represent the closer contracts better. Longer total distance means there are some contracts represented by the centroids further away. Dividing the points into subsets according to the space in these plots instead of dividing randomly may solve this problem, but it is difficult to divide the contracts with six different attributes in the same way, and the user-defined standards increase the risk of producing inaccurate results. For example, dividing the 400 points equally into two subsets by the x attributes is different with y attributes. The clustering results are different according to the decision made by the user.



**Fig. 4**

Given an extreme example, all centroids are in the centroid of the space if there is one centroid per subsets. It is obviously a poor clustering result. All in all, it produces a bad result with a small number of centroids per subset even we have a large number of total centroids.



Dup	100	500	1000	2000
1	76.88	419.05	827.42	1594.68
2	75.1	403.99	759.5	1237.13
5	5.96	41.06	95.51	237.28
10	2.58	44.08	126.82	317.93
50	2.41	37.27	92.84	202.97

Dup %	100	500	1000	2000
1	76.88	83.81	82.74	79.73
2	75.1	80.8	75.95	61.86
5	5.96	8.21	9.55	11.86
10	2.58	8.82	12.68	15.9
50	2.41	7.45	9.28	10.15

**Table.3**

The duplication and duplication in percentage of the MAS are presented. The column name is the total number of clusters. The row name is the number of cluster per subset.

$$\text{Dup \%} = \frac{\text{Dup}}{\text{the total number of clusters}} * 100$$

From Table 3, there are a lot of duplications happened in the MAS. The duplication decreases the number of representative contracts used to estimate the portfolio value in further steps, so it decreases the accuracy. There are three disadvantages of the duplication. First, we cannot control the desired number of distinct centroids after removing the duplication. Second, it would be a waste of time finding centroids that lead to duplicated representative contracts. Third, if there are no enough centroids due to duplication, it will produce a poor result. For example, there are 77 duplicated centroids in the 100 total centroids with 1 centroid per subset in Table 3. The accuracy of this example in Table 4 and Table 5 is poor because only 23 centroids are used in the MC and kriging steps.

Absolute Improvement	MAS				RDA				SRS			
	100	500	1000	2000	100	500	1000	2000	100	500	1000	2000
1	20.29	44.88	51.67	51.01	-14.6	-17.43	-12.38	-9.85	51.59	53.42	56.25	53.45
2	-2.16	15.75	27.52	18.52	-2.08	-9.24	-6.13	-3.92	3.4	16.08	26.44	17.99
5	-0.98	-1.73	0.22	1.55	2.74	-0.36	-1.78	-1.59	-2.86	-2.86	-1	-0.3
10	-1.02	-2.74	-1.94	0.22	0.62	0.36	0.28	0.99	-9.33	-3.85	-2.31	-0.76
50	0.05	-0.57	-0.65	0.00158	-0.03	0.04	-0.31	-0.15	-9.5	-3.28	-1.63	-0.57

**Table.4**

The absolute improvement of the difference are presented. The column name is the total number of clusters. The row name is the number of cluster per subset.

$$\text{Absolute Improvement} = (\text{Diff}_{MAS} - \text{Diff}_{other})$$

Relative Improvement	MAS				RDA				SRS			
	100	500	1000	2000	100	500	1000	2000	100	500	1000	2000
1	32	76	86	90	-23	-30	-21	-17	81	91	93	95
2	-14	74	91	89	-13	-43	-20	-19	22	75	87	86
5	-11	-78	10	61	31	-16	-79	-63	-32	-128	-45	-12
10	-36	-178	-142	11	22	24	21	50	-326	-251	-169	-39
50	3	-31	-33	0	-1	2	-16	-8	-479	-179	-83	-30

**Table.5**

The relative improvement of the difference are presented. The column name is the total number of clusters. The row name is the number of cluster per subset.

$$\text{Relative Improvement} = \frac{(\text{Diff}_{MAS} - \text{Diff}_{other})}{\text{Diff}_{MAS}}$$

Because of this, the following three alternative methods without duplication are tested and presented. From Table 4, the MWS method has a better performance on the upper right corner of Table 4 and Table 5, comparing with MAS. In the MAS method, if the centroids are very close to each other, one same contract may be selected as representative contract for two different centroids. The number of duplications ranges from 2 to 1600, or 2% to 80%, for different combinations of (total number of centroids) and (the number of centroids per subset). The negative impact on accuracy from the duplication is obvious. The reason why there is no duplication by the MWS method is that the 100000 synthetic contracts are different with the continuous uniform distribution of the premium attribute. The 100000 different synthetic contracts are divided into different subsets, so it is impossible to select same representative contracts from the different subset. From the simplified example in Figure 4, the centroids in the same subset are not close enough to have same representative contracts, so there is no duplication in the MWS method.

As the number of centroids per subset increase, we can see the duplication in the MAS method decrease and the accuracy is close to each other from the last row of Table 4, which is less than 1%. Also, it spends more time on the distance calculations as the number of centroids per

subset increase. However, by the MWS method, for each centroid, we only need to calculate the distances of the centroid to the contracts in the same subset to select the closest contract. By the MAS method, the distances of the centroid to the contracts in the whole portfolio are calculated to select one representative contract for the centroid. From the (MWS) and (time difference in percentage) part in Table 6, the MWS method saves a lot of time in step1comparing with the MAS method. Also, the MWS method saves a lot of time in (2,000\*50) case with less duplication in the MAS method.

In Table 4, for the RDA method, the result is better for the results on the bottom left corner of Table 4 and Table 5, comparing with MAS. There is no duplication in this method because centroids are mean of numerical values and most frequent category values of the attributes in contracts from this cluster. It is rare that contracts from two different clusters have the same mean. Even there is no duplication, the centroids, which have same representative contracts by the MAS method, approach to the same. Even it is not exact same, its performance is still very close to the MAS. Because of this, when the number of centroids per subset is small, it is not close to the real value, which is similar as MAS. However, it is better when the contracts are divided into subsets by an appropriate way and it has less influence from the centroids approach to the same. It has smallest total distance by using centroids directly. Additionally, this method is faster than the MAS method and MWS method because we skip the step of nearest mapping.

From Table 4 and Table 5, dividing the whole portfolio into different number of subsets does not influence the accuracy when we use the SRS method instead of clustering. It has a very good performance when there is a small number of centroids per subset, comparing with other three methods. We do not need to consider how to divide the subsets by the SRS method. The accuracy increases as the total number of sample increase. This method skips the clustering step and saves a lot of time. However, to get an accurate enough result, the sample size is larger than the other three methods, which increase the time of the Monte Carlo step and the Kriging

step. Further research is needed to test the trade-off between time of the clustering method and the time of the further two steps. If there are higher requirements for the accuracy, the clustering step is necessary to save time. Additionally, interesting objects may be observed by the clustering step. There is a conjecture that the uniform distributions of the attributes impact the accuracy. If a real large portfolio of VA contracts is priced, the clustering step may have more influence on the accuracy because it is not uniform distributed any more. The synthetic contracts are not good enough to test the accuracy of the algorithm.

100*1		Time	Time diff %	2000*50		Time	Time diff %
<b>MAS</b>				<b>MAS</b>			
	<b>step1</b>	2.66	0		<b>step1</b>	107.36	0
	<b>step2</b>	0.38	0		<b>step2</b>	34.52	0
	<b>step3</b>	0.32	0		<b>step3</b>	29.71	0
	<b>total</b>	3.37	0		<b>total</b>	168.19	0
<b>MWS</b>				<b>MWS</b>			
	<b>step1</b>	1.1	58.65		<b>step1</b>	75.55	29.63
	<b>step2</b>	1.78	-368.42		<b>step2</b>	34.29	0.67
	<b>step3</b>	1.69	-428.13		<b>step3</b>	33.46	-12.62
	<b>total</b>	4.57	-35.61		<b>total</b>	143.29	14.8
<b>RDA</b>				<b>RDA</b>			
	<b>step1</b>	1.15	56.77		<b>step1</b>	74.34	30.76
	<b>step2</b>	1.71	-350		<b>step2</b>	34.52	0
	<b>step3</b>	1.46	-356.25		<b>step3</b>	33.4	-12.42
	<b>total</b>	4.32	-28.19		<b>total</b>	142.25	15.42
<b>SRS</b>				<b>SRS</b>			
	<b>step1</b>	0.04	98.5		<b>step1</b>	0.06	99.94
	<b>step2</b>	1.87	-392.11		<b>step2</b>	34.54	-0.06
	<b>step3</b>	1.43	-346.88		<b>step3</b>	33.5	-12.76
	<b>total</b>	3.34	0.89		<b>total</b>	68.11	59.5

**Table.6**

The time (seconds) of the difference methods are presented. The different time in percentage comparing with the base line are presented. 100\*1 and 2000\*50 are the total number of centroids\* the number of centroids per subset. The row name is the three steps in Gan's paper.

In Table 6, 100\*1 by the MAS method spends less time on the Monte Carlo method and Kriging method because the duplicated centroids are removed. It spends around 350% to 400 % time for the step2 and step3 because the number of distinct centroids is around 23 in the MAS method, and  $100/23=435\%$ . Around 77 duplicated centroids are removed in the MAS method. However, the MAS method spends more time in the nearest contract mapping step. It saves 58.65%, 56.77% and 98.5% in the step1 for the MWS method, the Rounding method, and the SRS method. For the 2000\*50, the total time by the MAS method is greater than the total time by the MWS method obviously. It saves 14.8%, 15.42% and 59.5% in the total time for the MWS method, the Rounding method, and the SRS method. The time of the first step in the MWS method is even very close to the rounding method. The time on the Monte Carlo method and the Kriging method for the MWS method, the rounding method and the SRS method are close to each other because there is no duplication and the same number of representative contracts are calculated in these two steps.

**6 Conclusion**

Four different methods for identifying representative contracts in a large portfolio of VAs are alternatives of the three-step method proposed by Gan (2013). To compare the four different methods for identifying representative contracts in a large portfolio of VAs, we examine the four methods by comparing the accuracy and speed of the results. The MAS method is used by Gan (2013). By comparing MAS method and MWS method, MAS method could produce

duplication and increase the inaccuracy overall. By comparing MAS method and RDA method, we find that it is unnecessary to make nearest neighbor mapping. The MAS method, MWS method and RDA method include clustering, and the SRS method does not have clustering. For all three clustering methods simulated with comparable computation, we should perform cluster on large subsets rather than smaller. SRS is really fast in identifying representative contracts and its results are reasonably accurate.

In further research, we will consider a combination of SRS method and clustering method so that it has both high accuracy and fast. For example, we select a random sample and find total centroids from the random subset. This alternative method produces a good result when we tested the 5000 samples with 100 centroids. The mean of percentage error is 2.01%, comparing with the full-scaled Monte Carlo result and the mean of the time is 8.95 seconds. Secondly, more complicate VA benefits and stock dynamic can be tested. For example, we can consider guaranteed benefits with roll-up and ratchet. We can consider the stock dynamic such as regime-switching. Lastly, the substitutions of Monte Carlo method or the Kriging method also can be tested to find the preferred plan.

## Reference

- Ahmad, A., & Dey, L. (2007). A k-mean clustering algorithm for mixed numeric and categorical data. *Data & Knowledge Engineering*, 503-527.
- Bacinello, A. R., Millosovich, P., Olivieri, A., & Pitacco, E. (2011). Variable annuities: A unifying valuation approach. *Insurance: Mathematics and Economics*, 285-297.
- Bauer, D., Kling, A., & Russ, J. (2008). A universal pricing framework for guaranteed minimum benefits in variable annuities. *ASTIN Bulletin: The Journal of the IAA*, 621-651.
- Gan, G. (2013). Application of data clustering and machine learning in variable annuity valuation. *Insurance: Mathematics and Economics*, 795-801.
- ap, G., & Lin, X. S. . (2015). Valuation of large variable annuity portfolios under nested simulation: A functional data approach. *Insurance: Mathematics and Economics*, 138-150.
- Hardy, M. (2003). *Investment guarantees: modeling and risk management for equity-linked life insurance*. John Wiley & Sons.
- Huang, Z. (1998). Extensions to the k-means algorithm for clustering large data sets with categorical values. *Data mining and knowledge discovery*, 283-304.
- Isaaks, E. H. Srivastava RM. (1990). *An Introduction to Applied Geostatistics*. Oxford, UK: Oxford University Press.
- McNeil, A. J., Frey, R., & Embrechts, P. . (2015). *Quantitative risk management: Concepts, techniques and tools*. Princeton university press.
- Stanfill, C., & Waltz, D. (1986). Toward memory-based reasoning. *Communications of the ACM*, 1213-1228.