香港統計學會

Hong Kong Statistical Society

c/o Department of Statistics & Actuarial Science,
The University of Hong Kong, Pokfulam Road, Hong Kong
http://www.hkss.org.hk

Bulletin
Volume 38 No.1
March 2016

Editor's Foreword

As the new Editor of the Bulletin, I would like to thank my predecessor, Raymond TAM, for his admirable leadership during his tenure as the Editor. I also thank the members of the Editorial Board, Billy LI and Carly LAI for their valuable services to the Bulletin.

The Editorial Board has issued both the e-News and the Bulletin in 2015/16. They provide information about our activities, events, achievements of members on a timely basis. However, this can be done only if we can gain support from members. Therefore, I, on behalf of the Editorial Board, would like to invite members to contribute their works to our publications.

In this issue, among others, Chris LAI and Philip YAM wrote an introductory article on Mean Field Games which is a very hot topic in control theory and mathematical finance. Olivia OR and Agnes LAW contributed an article on the application of Big Data on official statistics.

I would like to invite members to contribute their works and share their research interests in this Bulletin so that we can promote our profession.

Ben CHAN

CONTENTS

(Vol. 38/No.1, March 2016)

President's Forum

Professor K.W. NG

The year 2015 marked a special year for the statistical community worldwide. In celebration of the second World Statistics Day on 20 October 2015, the Hong Kong Statistical Society (HKSS) jointly organised a public lecture with the Census and Statistics Department and the Education Bureau in the Lecture Hall of the Hong Kong Science Museum. Very favourable responses were received from the participants. Highlights of the public lecture are given in the News Section. The HKSS also signed up to a Statement for World Statistics Day launched by the Royal Statistical Society (RSS), the American Statistical Association and the International Statistical Institute to recognise the importance of statistics in the international development process. More details can be found in the News Section.

The Society is greatly honored by the organisers' invitations to be a sponsor of three conferences in 2016, namely the 2016 ICSA Conference to be held in China, the symposium on Frontier research in Statistics and Data Science to be held in Hong Kong, and the IMS Asia Pacific Rim Meeting (IMS-APRM) to be held in Hong Kong, which is the fourth meeting of the Institute of Mathematical Statistics meeting series. See the News Section for more details.

We congratulate more academics in probability and statistics in Hong Kong receiving official awards from China; see the News Section for details. Following the foot-steps of Professor Howell TONG of The University of Hong Kong in 2000 and Professor ZHU Li-xing of The Baptist University of Hong Kong in 2013 respectively, Professor SHAO Qi-man of The Chinese University of Hong Kong and Professor JING Bing-yi of The Hong Kong University of Science and Technology have jointly received the 2015 State Natural Science Award (Second Class) given by The State Council of China. And after Professor LAM Yeh in 2007 and Professor Tony W. K. FUNG in 2009, both of The University of Hong Kong, Professor ZHAO Xing-qiu of The Polytechnic University of Hong Kong has received the 2014 Natural Science Award (Second Class) of the Higher Education Outstanding Scientific Research Output Award given by The Ministry of Education of China. See the News Section for details.

Needless to say, we all look forward to the award recipients' granting interviews by HKSS to share their honors and joys, as did Professor ZHU Li-xing for the March 2014 issue of the Bulletin. If we have missed other awards which members have received, please let the Bulletin Editor of HKSS or the General Secretary know, so that we can share the good news.

In an article of this Bulletin, Professor Philip YAM and Mr. Chris LAI share with us some models in the Mean Field Games with applications -- a recently popular field of study in mathematics and statistics -- to understand this powerful tool.

Big Data contain a huge amount of information which provides a valuable source for better insight and decision making in nowadays society, but there are new challenges in proper application at the same time. Following two talks on the topic last year (cf. March 2015 Bulletin, News Section), we had three speakers from the software company SAS; see News Section in this Bulletin. We have also invited Ms. Olivia OR and Dr. Agnes LAW to share with us relevant issues in the use of Big Data and illustrate its usage in various official statistics such as price statistics, traffic and transport statistics and social media.

Thanks to the Organising Committees for the 2014/15 Statistical Project Competition for Secondary School Students under the leadership of Mr. Alan CHEUNG and the Statistics Creative-Writing Competition for 2014/15 under the leadership of Dr. Philip YU, the 2014/15 round of Competitions have been successfully concluded with details and photos presented in this Bulletin.

Following the decision of Royal Statistics Society to discontinue its professional examinations from September 2017, the HKSS Council has decided there will be no professional examinations after the 2017 round. We will keep members posted of the latest development in good time.

The Society had organised a day tour to Grass Island, Hoi Ha Wan Marine Park and Shui Long Wo on 13 March 2016. See the News Section for the highlights of this enjoyable social activity.

The 2015/16 Annual General Meeting (AGM) of the HKSS is scheduled to be held on **21 April 2016 (Thursday)** at 6:15pm at the Immigration Officer Mess, 20/F Immigration Tower in Wan Chai. I look forward to seeing you in the AGM!

What is a Mean Field Game?

C.H.C. Lai[†] and S.C.P. Yam[†]

[†]Department of Statistics, The Chinese University of Hong Kong

October 2, 2015

Modeling collective behavior is always of significant interest for scientists. In reality, large systems in which individuals interact with each other, known as interacting particle systems, are omnipresent and these systems usually possess complex structure, thus rendering the modeling task formidable. It can be easily imagined that the complexity of such a system increases with the number of individuals, so does the difficulty of its study.

As an example, on the second day of every Lunar New Year, hundreds of thousand citizens rush to the Star Avenue for good places to appreciate the firework. The shortest path is surely Nathan Road. If you were the only one going, you should definitely take Nathan Road. However, Hong Kong is really a crowded city. Alongside you are other spectators also looking for the fastest way to the Star Avenue. Thus, the decision process on determining the quick way is more complicated. If everyone still goes for Nathan Road, traffic congestion would slow them down. Therefore, some may use other routes, even of longer distances, which are less congested. Here the action of each citizen depends on the decision makings of other citizens. For example, they have to ponder over which routes the crowd would take in order to avoid congestion. This is an example of interacting particle system, where the particles are the firework spectators. In the following, particles, individuals and agents are interchangeable and they all stand for the participants in the system.

Interestingly, we are making decision taking into account actions of others in daily life. You may even think of examples like how you choose your transportation in daily life. In fact, interacting particle systems exist in statistical physics, finance, economics, sociology and other academic areas. Mean Field Games study these systems with a view to providing insight for prediction and forecast of the behavior of individuals inside the systems, and thus recently become one of the most popular fields of study in mathematics and statistics. In this article, we walk through the development of Mean Field Games with some toy models and applications, in order to understand this useful and powerful tool step by step.

1 Mean Field Theory

The development of Mean Field Games stems from statistical mechanics in physics. Consider a multiple-particle system, in general the interaction between particles can be of different kinds. The generality of interaction and high number of particles altogether make the system not explicitly tractable. Even though computers are fast and accurate nowadays, the computation time is extremely long and also increases exponentially with the number of particles.

In order to simplify the interaction, Pierre Weiss [20] first proposed the idea of Mean Field Theory in the study of phase transition of substances. Essentially, Mean Field Theory replaces all forms of interaction by the summarized average behavior of all particles. Because this approach greatly reduces the complexity of the system by imposing constraints on the type of interaction, applications have been seen later in epidemic studies, queuing theory, network problems and game theory.

For example, consider a system where N particles are on a straight line. The position of each particle is represented by a real number. For $i = 1, 2, \dots, N$, we have

$$X_{k+1}^i = c_k + A_k X_k^i + \bar{A}_k \left(\frac{1}{N-1} \sum_{j \neq i}^{N} X_k^j \right) + W_k^i, \text{ with } X_0^i = 0. \tag{1}$$

Here, c_k is a deterministic process. $X_k^i \in \mathbb{R}$ is the position of the i-th particle at time k and W_k^i's are some i.i.d. random noises with zero mean and finite variance. Here, we describe that as time evolves, the particle changes its position depending on its own previous position and the average position of all other particles. The third term, which summarizes the impact from other particles by the average position, is the origin of the term "Mean Field". The coefficients A_k and $\bar{A}_k$ measure the sensitivities of the new position towards the previous position and the average position of other particles respectively. The evolution of each particle is now more manageable because of this simpler structure of interaction. Indeed, the convenient property exhibited by this structure arises from the Law of Large Number. Here, all particles evolve according to the same structure, only with independent random noises. When N becomes large, i.e. N tends to infinity, the average position tends to the expected value of the position almost surely. The expected value is the same for any i because of symmetry, and hence we have

$$\frac{1}{N-1} \sum_{j \neq i}^{N} X_k^j \quad \xrightarrow{a.s.} \quad \mathbb{E}\big(X_k^i\big),$$

which implies

$$X_{k+1}^i = c_k + A_k X_k^i + \bar{A}_k \mathbb{E}(X_k^i) + W_k^i, \text{ with } X_0^i = 0. \tag{2}$$

Here, the expected value can be found easily. This provides mathematical tractability to the average behavior of the model. In particular, the Mean Field can be obtained by taking expectation on both sides of (2), which results in

$$\mathbb{E}\big(X_{k+1}^i\big) = c_k + A_k \mathbb{E}\big(X_k^i\big) + \bar{A}_k \mathbb{E}\big(X_k^i\big) = c_k + (A_k + \bar{A}_k)\mathbb{E}(X_k^i).$$

Now, the position of the i-th particles does not explicitly depend on any other individual particle, but on the community as a whole. Obviously, it can be calculated easily with $A_k + \bar{A}_k \neq 1$ by solving a linear equation. Conceptually, this can be understood in the following way. Originally, the position is affected by the other $N-1$ particles and all these particles are positioned differently because of their independent noises. However, when N is very large, the uncertainty of the noises are averaged out, giving rise to the nice property aforementioned.

The above toy model demonstrates a case where particles evolve freely without any disruption. In economics, biology and other fields, it is often that a control is imposed at one's will to alter the behavior of his own particle in order to adapt to the environment. This results in the mathematical concepts of optimization and games. In fact, when Mean Field Theory is combined with Stochastic Differential Games, we obtain a powerful tool to explain interacting particle systems naturally existing in diverse disciplines.

2 Stochastic Differential Games (SDGs)

In many single-agent problems, the agent is required to choose the most suitable action from a set of possible actions to achieve a certain goal. Translated to mathematical languages, this means the agent has to choose an optimal control, usually denoted by u, from an admissible control set $\mathcal{A}$, in order to optimize a certain cost or gain functional J. The process is known as optimization. The cost functional might be traveling time, operation cost or any other quantifiable entries while the respective admissible sets are the possible transport routes, feasible actions subject to capital restriction or other possible constraints.

Yet, in the real world, rarely do the settings consist only one agent. In biology, economics, finance, sociology and networking problems, usually multiple agents are involved. Recall our firework example which is a multiple-agent situation. If optimization is adopted independently by each spectator, it is obvious that they should all choose Nathan Road. Nevertheless, these decisions altogether definitely will not work because of the limited road capacity. Therefore, a different mathematical idea has to be deployed instead of optimization, by taking the decision makings of all agents into account.

Game theory is developed to handle this situation especially in the non-cooperative setting, in which each agent may not maintain a perfect compromise with the rest. In a non-cooperative game, or a game for short in the rest of this article, every agent seeks to make the best decision, taking into account all possible actions of others. The resulting scenario is known as the Nash Equilibrium, first proposed by Nobel Prize laureate John Nash [18] and other followers. Formally, for a game in which the goal is to reduce cost, the Nash Equilibrium is defined as follows.

Definition 2.1. *For an N-player game with the admissible action set $\mathcal{A}$ and cost functional $J(v_1, \dots, v_N)$, if the action $u = (u_1, u_2, \dots, u_N)$, satisfies, for all i,*

$$J(u_1, \dots, u_{i-1}, u_i, u_{i+1}, \dots, u_N) \leq J(u_1, \dots, u_{i-1}, v_i, u_{i+1}, \dots, u_N),$$

where $v_i \in \mathcal{A}$ is any admissible action. Then the action u constitutes a Nash Equilibrium for the game.

In other words, each agent has no intention to switch from his own equilibrium strategy because he would be worse off if all the others stick with the equilibrium. In our firework example, spectators search for different routes, with traveling time as their cost. Eventually, they are distributed in different routes so that the traveling times on each route become similar. Now, if a spectator alters his route from this equilibrium, they would actually be slower because of the extra traffic burden to the route that he changes to. Therefore, they would abide by the equilibrium.

To complete the setting of a Stochastic Differential Game, we also have to model the state which is to be altered by the control in a dynamical way. The most common way to describe the random evolution of state mathematically is via Stochastic Differential Equations, thus giving rise to the name Stochastic Differential Games. Here, as a motivation, we give a discrete time analog of it, which is essentially an extension of (1) by imposing a control. Interested readers can refer to the Appendix for the continuous time version.

$$X_{k+1}^i = c_k + A_k X_k^i + \bar{A}_k \left(\frac{1}{N-1} \sum_{j \neq i}^{N} X_k^{j^*} \right) + B_k v_k^i + W_k^i, \text{ with } X_0^i = 0, \tag{3}$$

where v_k is the control on the i-th particle imposed at time k to adjust its position at time $k+1$. Here, other agents are imposing the equilibrium strategy, thus their states are denoted by $X_k^{j^*}$, and we look for the control that the i-th agent would adopt under this situation. When N increases, in general it gets harder to obtain an explicit and mathematically tractable solution for the above case. But thanks to the Law of Large Number, when N tends to infinity, the equation becomes

$$X_{k+1}^i = c_k + A_k X_k^i + \bar{A}_k \mathbb{E}\big(X_k^{i^*}\big) + B_k v_k^i + W_k^i, \text{ with } X_0^i = 0,$$

where the Mean Field can be first obtained. Thus, the problem can be tackled by the notion of Mean Field Games.

Over the past few decades, the study of SDGs has been a major research topic in control theory and financial economics, especially in studying the continuous-time decision making problem between non-cooperative investors. In regard to the one-dimensional setting the theory of two person zero-sum games is quite well-developed via the notion of viscosity solutions, see for example Elliott [10], and Fleming and Souganidis [11]. Unfortunately, most interesting SDGs are N-player non-zero sum ones. In this direction, interested readers are referred to the works of Bensoussan and Frehse [1, 4] and Bensoussan et al. [5], but there are still relatively few results in the literature mainly because of the underlying complicated mathematical structure.

3 Mean Field Games (MFGs)

The development of Mean Field Games can be dated back to 2003, when Huang et. al. [13] first studied stochastic differential games involving infinitely many players. Their problem originated from wireless communication network where they applied a macroscopic view to look for the Nash Equilibrium, under the name "Large Population Stochastic Dynamics Games". Independently, Lasry and Lions [15, 16, 17] adopted Mean Field Theory for the interacting systems and solved for the equilibrium. Therefore, it gives rise to the name "Mean Field Game".

Inherited from Mean Field Theory, the Mean Field term serves as a summary of the community about the actions that all other agents would take. Given the Mean Field, a specific agent, i.e. the i-th agent, aims to optimize his own objective functional. Under the homogeneity assumption, it turns out the action that the i-th agent solves for would be the action that the whole community should take, because it is the best possible strategy, assuming all other agents are intelligent. Therefore, we arrive at an equilibrium.

In Mean Field Games, there are assumptions that follow from Mean Field Theory. First, there are infinitely many agents of homogeneous dynamics and objectives. This is to ensure that action of a single agent has negligible effect on the community consent, which is the Mean Field. Second, interaction is conducted only through the Mean Field. Therefore, the interaction is manageable and the Mean Field summarizes the action of the community.

As an example, we investigate here a discrete time version of Linear-Quadratic MFGs in Bensoussan et al. [8] which gives an explicit and tractable solution. The original continuous time version in [8] can be found in Appendix.

As an illustration for this complicated concept, consider a 2-period game. The agent aims to minimize the following cost functional,

$$J(v) = \sum_{k=0}^{1} \mathbb{E}\Big[Q_k {X_k^v}^2 + R_k v_k^2 + \bar{Q}_k \big(X_k^v - S_k \mathbb{E}(Y_k)\big)^2\Big] + \mathbb{E}\Big[Q_2 {X_k^v}^2 + \bar{Q}_2 \big(X_2^v - S_2 \mathbb{E}(Y_2)\big)^2\Big],$$

where $R_k, Q_k, \bar{Q}_k > 0$. Here, the first summation is the running cost from time 0 to 1 and the second term is the terminal cost. The terms $Q_k {X_k^v}^2$ is the quadratic cost involved. Also, agents are penalized by $\bar{Q}_k\big(X_k^v - S_k\mathbb{E}(Y_k)\big)^2$ if their state deviates from the community consent. The cost of imposing control is captured by $R_k v_k^2$. The state is given by, for $k = 0, 1$,

$$\begin{cases} X_{k+1}^v &= A_k X_k^v + \bar{A}_k \mathbb{E}(Y_k) + B_k v_k + W_k, \\ X_0^v &= x_0. \end{cases} \tag{4}$$

In the above, v_k is the control by the agent to alter the state X_k^v, where the superscript indicates the control applied. Y_k is the optimal state corresponding to the optimal control u_k of this problem. The Mean Field term $\mathbb{E}(Y_k)$ is the the summary of the community state. Under Nash's philosophy, it is reasonable that the social consent advocates the equilibrium strategies u_k as it minimizes the cost. Therefore, the agent should aim to minimize his own cost given the preferences of the others.

3.1 Establishing a Necessary Condition

At each time k, the agent observes its state X_k^v and makes the decision on v_k accordingly. Now, we consider a perturbed control $u + \theta v$ around the optimal control u, where θ is any real number and v is any control. First, notice that the state $X_k^{u+\theta v}$ changes if θ changes. To explicitly express the relationship between X_k and θ, we see that the perturbed state is given by

$$\begin{cases} X_{k+1}^{u+\theta v} &= A_k X_k^{u+\theta v} + \bar{A}_k \mathbb{E}(Y_k) + B_k(u_k + \theta v_k) + W_k, \\ X_0^{u+\theta v} &= x_0. \end{cases} \tag{5}$$

Now, define a process $\tilde{X}_k^v$ as follows.

$$\begin{cases} \tilde{X}_{k+1}^v &= A_k \tilde{X}_k^v + B_k v_k, \\ \tilde{X}_0 &= 0. \end{cases} \tag{6}$$

Then, we consider the linear combination as below. For $k = 0, 1$,

$$\begin{aligned} X_{k+1}^u + \theta \tilde{X}_{k+1}^v &= A_k X_k^u + \bar{A}_k \mathbb{E}(Y_k) + B_k u_k + W_k + \theta A_k \tilde{X}_k^v + \theta B_k v_k \\ &= A_k (X_k^u + \theta \tilde{X}_k^v) + \bar{A}_k \mathbb{E}(Y_k) + B_k (u_k + \theta v_k) + W_k. \end{aligned}$$

Comparing this with (5), we establish the linear relationship

$$X_k^{u+\theta v} = X_k^u + \theta \tilde{X}_k^v. \tag{7}$$

As u is the optimal control, the minimum cost is attained when θ equals 0. Therefore, the derivative with respect to θ at that point is zero. This gives the first order condition:

$$\begin{aligned}0 &= \frac{d}{d\theta}\bigg|_{\theta=0} J(u+\theta v)\\ &= \frac{d}{d\theta}\bigg|_{\theta=0}\left\{\sum_{k=0}^{1}\mathbb{E}\Big[Q_k {X_k^{u+\theta v}}^2 + R_k(u_k+\theta v_k)^2 + \bar{Q}_k\big(X_k^{u+\theta v} - S_k\mathbb{E}(Y_k)\big)^2\Big] + \mathbb{E}\Big[Q_2 {X_2^{u+\theta v}}^2 + \bar{Q}_2\big(X_2 - S_2\mathbb{E}(Y_2)\big)^2\Big]\right\},\end{aligned}$$

which is equivalent to the following, by applying the linear relationship (7) and differentiating,

$$2\mathbb{E}\Big[v_0R_0u_0 + \tilde{X}_1(Q_1+\bar{Q}_1)^v X_1^u + v_1R_1u_1 - \tilde{X}_1^v\bar{Q}_1S_1\mathbb{E}(Y_1)\Big] + 2\mathbb{E}\Big[\tilde{X}_2^v(Q_2+\bar{Q}_2)X_2^u - \tilde{X}_2^v\bar{Q}_2S_2\mathbb{E}(Y_2)\Big] = 0.$$

Without causing ambiguity and for simplicity, we use X_k and $\tilde{X}_k$ in place of X_k^u and $\tilde{X}_k^v$ respectively. By rewriting the subscript of the controls and using summation to represent them, the above necessary condition can be rewritten as

$$\mathbb{E}\left[\sum_{k=1}^{2}\Big(v_{k-1}R_{k-1}u_{k-1} + \tilde{X}_k(Q_k+\bar{Q}_k)X_k - \tilde{X}_k\bar{Q}_kS_k\mathbb{E}(Y_k)\Big)\right] = 0. \tag{8}$$

3.2 Constructing the Optimal Control

Therefore, we look for a control u which can satisfy (8) for arbitrary control v. To achieve so, we adopt the Adjoint Equation Approach. Put it simple, we construct an adjoint process according to the cost structure so that (8) is satisfied when the control is expressed in terms of the adjoint process. In particular, we construct the following adjoint process.

$$\begin{cases} p_2 &= (Q_2+\bar{Q}_2)X_2 - \bar{Q}_2S_2\mathbb{E}(Y_2),\\ p_1 &= A_1p_2 + (Q_1+\bar{Q}_1)X_1 - \bar{Q}_1S_1\mathbb{E}(Y_1) - A_1\big(p_2 - \mathbb{E}(p_2|X_1,X_0)\big).\end{cases} \tag{9}$$

Then, we have the following difference by telescoping terms:

$$\begin{aligned}\tilde{X}_1p_2 &= \tilde{X}_1(p_2-p_1) + \tilde{X}_1p_1\\ &= \tilde{X}_1(1-A_1)p_2 + v_0B_0p_1 - \tilde{X}_1(Q_1+\bar{Q}_1)X_1 + \tilde{X}_1\bar{Q}_1S_1\mathbb{E}(Y_1) + \tilde{X}_1A_1\big(p_2 - \mathbb{E}(p_2|X_1,X_0)\big).\end{aligned} \tag{10}$$

Here, for the term $\tilde{X}_1(p_2-p_1)$, we substitute p_1 by its definition in (6). For the second term $\tilde{X}_1p_1$, we use the definition of $\tilde{X}_1$ in (9). Similarly, we have

$$\begin{aligned}-\tilde{X}_1p_2 &= \tilde{X}_2p_2 + (\tilde{X}_2 - \tilde{X}_1)p_2\\ &= -\tilde{X}_1(1-A_1)p_2 + v_1B_1p_2 - \tilde{X}_2(Q_2+\bar{Q}_2)X_2 + \tilde{X}_2\bar{Q}_2S_2\mathbb{E}(Y_2),\end{aligned} \tag{11}$$

where the definitions of p_2 and $\tilde{X}_2$ are used in the first term and second term respectively. Summing the above two equations (10) and (11), and taking expectation, we arrive at

$$\mathbb{E}\left[\sum_{k=1}^{2}\Big(\tilde{X}_k(Q_k+\bar{Q}_k)X_k - \tilde{X}_k\bar{Q}_kS_k\mathbb{E}(Y_k)\Big)\right] = \mathbb{E}\left[\sum_{k=1}^{2}v_{k-1}B_{k-1}p_k\right], \tag{12}$$

where the difference between p_2 and its conditional expectation disappears by the Tower Property as

$$\begin{aligned}\mathbb{E}\Big[\tilde{X}_1A_1\big(p_2 - \mathbb{E}(p_2|X_1,X_0)\big)\Big] &= \mathbb{E}\Big[\mathbb{E}\Big[\tilde{X}_1A_1\big(p_2 - \mathbb{E}(p_2|X_1,X_0)\big)\Big|X_1,X_0\Big]\Big]\\ &= \mathbb{E}\Big[\tilde{X}_1A_1\mathbb{E}\Big[p_2 - \mathbb{E}(p_2|X_1,X_0)\Big|X_1,X_0\Big]\Big]\\ &= \mathbb{E}\Big[\tilde{X}_1A_1\big(\mathbb{E}(p_2|X_1,X_0) - \mathbb{E}(p_2|X_1,X_0)\big)\Big]\\ &= 0.\end{aligned}$$

Now, we are ready to establish the finalized necessary condition. By substituting (12) into (8), we have

$$\mathbb{E}\left[\sum_{k=1}^{2} v_{k-1}(R_{k-1}u_{k-1} + B_{k-1}p_k)\right] = 0. \tag{13}$$

Since v is arbitrary, it is very tempting to conclude that $u_k = -R_k^{-1}B_k p_{k+1}$. Nevertheless, when the agent makes decision at time k, he only observes his state up to that moment. However, p_{k+1} involves X_{k+1} so it cannot be used at time k. The correct optimal solution is obtained by taking a conditional expectation, which is

$$u_0 = -R_0^{-1}B_0\mathbb{E}(p_1|X_0) \quad \text{and} \quad u_1 = -R_1^{-1}B_1\mathbb{E}(p_2|X_0, X_1),$$

which can be obtained by further applying Tower property to (13).

To summarize, the idea of looking for the optimal control here is to first establish a necessary condition by perturbing the control. Then, we construct an adjoint process in terms of which the optimal control is defined so that the necessary condition is satisfied. Moreover, some readers may wonder if the optimal control obtained is also sufficient. The answer is yes. Because the cost functional is quadratic, it is a convex and coercive objective so that the optimal control is unique. A simple analog to this concept is that a strictly convex function would only have one single minimum point.

3.3 Linear Feedback Form of Optimal Control

In fact, a nice property of linear-quadratic control problem is that the control can be written into a feedback form. A feedback control is a control where it can be directly obtained from the contemporary observations without tracing the full history. In particular, if we take conditional expectation on p_{k+1} according to (9), by using the definition of X and the fact that the noises are independent, we can obtain the linear forms easily. Take p_2 as an example. By taking conditional expectation given the observation X_0 and X_1, we have

$$\mathbb{E}(p_2|X_0, X_1) = (Q_2 + \bar{Q}_2)\mathbb{E}(X_2|X_0, X_1) - \bar{Q}_2 S_2 \mathbb{E}(Y_2). \tag{14}$$

Now, by the definition of X_2 in (4) under the optimal control u, taking conditional expectation again yields that

$$\mathbb{E}(X_2|X_0, X_1) = A_1X_1 + \bar{A}_1\mathbb{E}(Y_1) - B_1^2R_1^{-1}\mathbb{E}(p_2|X_0, X_1).$$

The noise W_1 is independent of W_0 and x_0, so it is also independent of X_1 and X_0. Therefore it vanishes because it is of mean zero. Putting the above result back into (14) and by simple algebra, we obtain $\mathbb{E}(p_2|X_0, X_1)$ in terms of X_1 and some deterministic constants. Thus, we have u_1. u_0 is obtained similarly by calculating $\mathbb{E}(p_1|X_0)$. Interested readers can try as an exercise. We provide the result below.

$$\begin{cases} u_1 &= -\dfrac{B_1}{R_1C}\big[KX_1 + M_1\mathbb{E}(Y_1) + M_2\mathbb{E}(Y_2)\big], \\ u_0 &= \dfrac{B_0}{R_0 + B_0^2\left(Q_1 + \bar{Q}_1 + \frac{A_1K}{C}\right)}\left[A_0\left(Q_1 + \bar{Q}_1\dfrac{A_1K}{C}\right)X_0 + \bar{A}_0\left(Q_1 + \bar{Q}_1\dfrac{A_1K}{C}\right)\mathbb{E}(Y_0)\right. \\ & \left. \quad + \left(\dfrac{A_1M_1}{C} - \bar{Q}_1S_1\right)\mathbb{E}(Y_1) - \dfrac{A_1M_2}{C}\right], \end{cases} \tag{15}$$

where the coefficients are given by

$$C = 1 + \frac{B_1^2(Q_2 + \bar{Q}_2)}{R_1}, \quad K = A_1(Q_2 + \bar{Q}_2), \quad M_1 = \bar{A}_1(Q_2 + \bar{Q}_2) \quad \text{and} \quad M_2 = \bar{Q}_2S_2$$

The optimal control may seem a bit complicated but in fact it is readily applicable. It is in a linear form of the state and mean-field, with deterministic coefficients. The Mean Field term $\mathbb{E}(Y_k)$ can be calculated in advance by plugging these controls into the state and taking expectation. Then, with all the coefficients computed, the agent simply observes the state in each state and calculate the control value according to the linear formula given above. This type of control is known as linear feedback control.

3.4 Relationship with High Population Game

As demonstrated, MFGs provides mathematical tractability in microscopic level by a computable control. Interaction is also well taken into account through the Mean Field term as a social consent of the community in macroscopic structure. In application, the above Mean Field Game is a limiting case of a high-population game, when the number of agents tend to infinity. In particular, the optimal control (15) above serves as an approximation to the equilibrium strategy of the high-population game, which is stated as following.

Problem 3.1. *Consider N agents, each of whom control a state as in* (3)*, with $c_k = 0$, i.e.*

$$X^i_{k+1} = A_k X^i_k + B_k v^i_k + \bar{A}_k \left(\frac{1}{N-1} \sum_{j \neq i}^{N} X^{j^*}_k \right) + W^i_k, \ \ \textit{with } X^i_0 = 0,$$

and the cost functional is

$$J(v) = \sum_{k=0}^{1} \mathbb{E} \left[Q_k X_k^2 + R_k v_k^2 + \bar{Q}_k \left(X_k - S_k \frac{1}{N-1} \sum_{j \neq i} X^{j^*}_k \right)^2 \right] \\ + \mathbb{E} \left[Q_2 X_2^2 + \bar{Q}_2 \left(X_2 - S_2 \frac{1}{N-1} \sum_{j \neq i} X^{j^*}_k \right)^2 \right].$$

Find the control u which gives the minimum cost where $X^{j^*}_k$'s are the state of other agents with $i \neq j$ who also adopt u.

The above high-population problem can be extremely difficult to solve even only with a slight modification, given its high number of agents and complexity because of interaction. However, via MFG, we can find a solution which is an approximation of the Nash equilibrium of this game. And the approximation gets very close to the real solution even with only a few thousand agents.

For more advanced example and mathematical illustration, readers can refer to Appendix and the references therein. In conclusion, Mean Field Games is a useful tool to tackle problems of high-population interacting system as demonstrated. It provides an good approximation of the complicated system and allows us to retain the interaction while simplifying it by Law of Large Number. The Mean Field term is plausible in reality as a social consent and also provides mathematical tractability. Application is extensive as in economics, biology, network problems, etc.

4 Appendix

In SDGs, one of the first studied setting is the two-player zero-sum game proposed by Fleming and Souganidis [11]. Two players exert their control on a common state and share a common cost functional. One will aim to maximize while the other aim to minimize. The state follows the following Stochastic Differential Equation,

$$dX_t = f(t, X_t, Y_t, Z_t)dt + \sigma(t, X_t, Y_t, Z_t)dW_t,$$

where the function f and σ satisfy certain continuity and growth requirements to ensure the unique existence of the process X_t, and W_t here is now a Brownian motion. The cost functional is

$$J(Y, Z) = \mathbb{E}\Big[\int_0^T h(t, X_t, Y_t, Z_t)dt + g(X_T) \Big].$$

In this game, the state is X_t. Player 1 controls Y_t and try to minimize the cost functional J over all possible Z_t while Player 2 controls Z to maximize the cost over all Y_t.

The existence of equilibrium for this setting is provided in [11]. Interested readers can refer to the paper for detailed derivations. In fact, the holistic theory of Stochastic Differential Equation is used to model the states in many games, if not all.

Here, we also give the original continuous-time setting of the simplified version we discussed in Section 3 as in Bensoussan et al. [8]. The state follows a linear stochastic differential equation and the cost is quadratic, giving rise to the name of Linear-Quadratic MFGs. Linear quadratic set-up is studied because they often provide explicit and neat solution which are very tractable, while it happens naturally in diverse disciplines. Thus, it forms an initial step to understand the situation. In particular, the problem is given below.

Problem 4.1. *Let W_t be a Brownian motion and x_0 be a random variable with finite variance, independent of W_t. Find an square-integrable equilibrium strategy $u_t \in \mathbb{R}^m$ which minimizes*

$$J(v) = \mathbb{E}\left[\frac{1}{2}\int_0^T X_t^* Q_t X_t + v_t^* R_t v_t + (X_t - S_t Z_t)^* \bar{Q}_t (X_t - S_t Z_t) dt\right] + \mathbb{E}\left[\frac{1}{2} X_T^* Q_T X_T + (X_T - S_T Z_T)^* \bar{Q}_T (X_T - S_T Z_T)\right], \quad (16)$$

where the state in $\mathbb{R}^n$ follows

$$\begin{cases} dX_t &= \big(A_t X_t + B_t v_t + \bar{A}_t E(Y_t)\big) dt + \sigma_t dW_t, \\ X_0 &= x_0. \end{cases} \quad (17)$$

Here, v_t is the alterable control and Y_t is the optimal state corresponding to the equilibrium strategy u_t. All matrix coefficients are bounded and Q_t, R_t and $\bar{Q}_t$ are positive definite.

The Mean Field term $\mathbb{E}(Y_t)$ is the summary of the community state. As in the discrete case, the social consent advocates the equilibrium strategies u_t as it minimizes the cost. Therefore, the agent should aim to minimize his own cost under this scenario. Upon solving the game, the optimal control is $-R_t^{-1} B_t^* p_t$ where optimal state and strategy satisfy the following Forward-Backward Stochastic Differential Equation system.

$$\begin{cases} dY_t &= \big(A_t Y_t - B_t R_t^{-1} B_t^* p_t + \bar{A}_t \mathbb{E}(Y_t)\big) dt + \sigma_t dW_t. \\ Y_0 &= x_0, \\ -dp_t &= \big(A_t^* p_t + (Q_t + \bar{Q}_t) Y_t - \bar{Q}_t S_t \mathbb{E}(Y_t)\big) dt - Z_t dW_t, \\ p_T &= (Q_T + \bar{Q}_T) Y_T - \bar{Q}_T S_T \mathbb{E}(Y_T). \end{cases} \quad (18)$$

From the above system, we see that the forward equation describes the evolution of the state under the optimal strategy, while the backward equation traces the optimal strategy according to the cost structure from the terminal cost. Readers may observe that this system resembles its discrete counterpart in terms of structure. Similar as in the discrete case, another nice property of Linear Quadratic MFGs is that the optimal strategy can be written into a feedback form of the state $-R^{-1} B^* (\Xi y + \zeta)$, where Ξ and ζ satisfy

$$\begin{cases} \dfrac{d\Xi_t}{dt} + \Xi_t A_t + A_t^* \Xi_t - \Xi_t (B_t R_t^{-1} B_t^*) \Xi_t + Q_t + \bar{Q}_t = 0, \\ \Xi_T = Q_T + \bar{Q}_T, \\ \dfrac{d\zeta_t}{dt} = -A_t^* \zeta_t + \Xi_t (B_t R_t^{-1} B_t^*) \zeta_t + (\bar{Q}_t S_t - \Xi_t \bar{A}_t) z_t, \\ \zeta_T = -\bar{Q}_T S_T z_T. \end{cases}$$

The continuous MFG is also applied to approximate a high-population game, which is in fact the continuous counterpart of Problem 3. Here, we give the definition of the game.

Problem 4.2. *Let W_t^i for $i = 1, \ldots, N$ be N independent Brownian motions and x_0^i for $i = 1, \ldots, N$ be N independent finite-variance random variables also independent of the Brownian motions. Consider an N-player game, the state of the i-agent is given by*

$$\begin{cases} dx_t^i &= \left(A_t x_t^i + B_t v_t^i + \bar{A}_t \cdot \dfrac{1}{N-1} \displaystyle\sum_{j=1, j\neq i}^{N} x_t^j\right) dt + \sigma_t \, dW_t^i, \\ x^i(0) &= x_0^i, \end{cases} \quad (19)$$

and the cost functional is

$$
\begin{aligned}
&\mathcal{J}^i(v^1,\ldots,v^N)\\
&=\mathbb{E}\left[\frac{1}{2}\int_0^T (x_t^i)^* Q_t x_t^i + (v_t^i)^* R_t v_t^i\,dt + \frac{1}{2}(x_T^i)^* Q_T x_T^i\right]\\
&\quad+\mathbb{E}\left[\frac{1}{2}\int_0^T \left(x_t^i - S_t\cdot\frac{1}{N-1}\sum_{j=1,j\neq i}^N x_t^j\right)^* \bar{Q}_t \left(x_t^i - S_t\cdot\frac{1}{N-1}\sum_{j=1,j\neq i}^N x_t^j\right) dt\right]\\
&\quad+\mathbb{E}\left[\frac{1}{2}\left(x_T^i - S_T\cdot\frac{1}{N-1}\sum_{j=1,j\neq i}^N x_T^j\right)^* \bar{Q}_T \left(x_T^i - S_T\cdot\frac{1}{N-1}\sum_{j=1,j\neq i}^N x_T^j\right)\right].
\end{aligned}
$$

Find the control set $(u^1,\ldots,u^N)$ *which gives the Nash Equilibrium.*

For more details on the approximation and the relationship between MFGs and high-population systems, interested readers can refer to Bensoussan et al. [8] and Huang et al. [14]. Some other generalizations can be found in Bensoussan et. al. [2, 3].

The general setting of MFGs is more complicated. The forward-backward structure therefore consists of the Forward Fokker-Planck (FP) Equation, which describes the evolution of the state distribution under the optimal strategy, and the Hamilton-Jacobi-Bellman (HJB) equation, which describes implicitly the equilibrium control. The resulting system is known as FP-HJB system. For more details, we mentioned the blog of Tao [19], the book of Bensoussan et. al. [6], Bensoussan et. al. [7], Cardaliaguet [9] and Guéant et al [12].

Because of the extensive application, the literature of MFGs is rich. More general models have been developed. Mathematicians also analytically investigate the existence and uniqueness of MFGs by both partial differential equation approach and probabilistic approach. Interested readers are referred to the references and the citations therein.

References

[1] A Bensoussan and J Frehse. Stochastic games for n players. *Journal of Optimization Theory and Applications*, 105(3):543–565, 2000.

[2] Alain Bensoussan, Michael Man Ho Chau, and Sheung Chi Phillip Yam. Mean field games with a dominating player. *Applied Mathematics and Optimization*, 2015.

[3] Alain Bensoussan, Michael Man Ho Chau, and Sheung Chi Phillip Yam. Mean field stackelberg games: Aggregation of delayed instructions. *SIAM Journal on Control and Optimization*, 53(4):2237–2266, 2015.

[4] Alain Bensoussan and Jens Frehse. *On Diagonal Elliptic and Parabolic Systems with Super Quadratic Hamiltonians*, volume 8. 2009.

[5] Alain Bensoussan, Jens Frehse, and Jens Vogelgesang. Systems of bellman equations to stochastic differential games with non-compact coupling. *Discrete and Continuous Dynamical Systems*, 27(4):1375–1389, 2010.

[6] Alain Bensoussan, Jens Frehse, and Sheung Chi Phillip Yam. *Mean field games and mean field type control theory*. Springer, 2013.

[7] Alain Bensoussan, Jens Frehse, and Sheung Chi Phillip Yam. The master equation in mean field theory. *Journal de Mathématiques Pures et Appliquées*, 103(6):1441–1474, 2015.

[8] Alain Bensoussan, Joseph Sung, Sheung Chi Phillip Yam, and Siu Pang Yung. Linear-quadratic mean field games. *Submitted to Journal of Optimization Theory and Applications*, 2014.

[9] Pierre Cardaliaguet. Notes on mean field games. Technical report, 2010.

[10] Robert Elliott. The existence of value in stochastic differential games. *SIAM Journal on Control and Optimization*, 14(1):85–94, 1976.

[11] Wendell H Fleming and Panagiotis E Souganidis. On the existence of value functions of two-player, zero-sum stochastic differential games. *Indiana University Mathematics Journal*, (38), 1989.

[12] Olivier Guéant, Jean-Michel Lasry, and Pierre-Louis Lions. Mean field games and applications. In *Paris-Princeton lectures on mathematical finance 2010*, pages 205–266. Springer, 2011.

[13] Minyi Huang, Peter E Caines, and Roland P Malhamé. Individual and mass behaviour in large population stochastic wireless power control problems: centralized and nash equilibrium solutions. In *Decision and Control, 2003. Proceedings. 42nd IEEE Conference on*, volume 1, pages 98–103. IEEE, 2003.

[14] Minyi Huang, Roland P Malhamé, and Peter E Caines. Large population stochastic dynamic games: closed-loop mckean-vlasov systems and the nash certainty equivalence principle. *Communications in Information & Systems*, 6(3):221–252, 2006.

[15] Jean-Michel Lasry and Pierre-Louis Lions. Jeux à champ moyen. i–le cas stationnaire. *Comptes Rendus Mathématique*, 343(9):619–625, 2006.

[16] Jean-Michel Lasry and Pierre-Louis Lions. Jeux à champ moyen. ii–horizon fini et contrôle optimal. *Comptes Rendus Mathématique*, 343(10):679–684, 2006.

[17] Jean-Michel Lasry and Pierre-Louis Lions. Mean field games. *Japanese Journal of Mathematics*, 2(1):229–260, 2007.

[18] John Nash. Non-cooperative games. *Annals of Mathematics*, pages 286–295, 1951.

[19] Terrence Tao. Mean field games. *https://terrytao.wordpress.com/2010/01/07/mean-field-equations/*, January 23, 2011.

[20] Pierre Weiss. L'hypothèse du champ moléculaire et la propriété ferromagnétique. *J. Phys. Theor. Appl.*, 6(1):661–690, 1907.

Big Data and its Application in Official Statistics

Ms. Olivia OR and Dr. Agnes LAW

Introduction

As information and communications technology (ICT) continues to develop, the amount of data generated keeps increasing every day. This sheer amount of data, called Big Data, is defined by its characteristics along three dimensions (3Vs - high **v**olume, **v**elocity and **v**ariety). Volume is the number of data records, their attributes and linkages. Velocity is how fast the data are produced and changed, and the speed at which they must be received, processed and understood. Variety is the diversity of data sources, formats, media and content. Most of the time, Big Data is generated in digital form, and is unstructured and complex. Thus, new forms of processing and data management beyond the traditional tools are required.

Big Data contains a huge amount of information for better insight and decision making. Because of its enormous value, companies and researchers are eager to develop new technologies and methodologies to extract and analyse the information from Big Data. Realising that Big Data is a potential resource for better decision making, some National Statistical Offices (NSOs) and Government agencies around the world have experimented with adopting Big Data in official statistics.

Common sources of Big Data

As stated by the United Nations Economic and Social Council (2014), the common sources of Big Data can be broadly classified as follows:

- Sources arising from the administration of a programme, be it governmental or not, e.g. electronic medical records, insurance records, bank records and food banks;
- Commercial or transactional sources arising from the transactions between two entities, e.g. credit card transactions and online transactions (including those from mobile devices);
- Sensor network sources, e.g. satellite imaging, road sensors and climate sensors;
- Tracking device sources, e.g. tracking data from mobile telephones and the Global Positioning System (GPS);
- Behavioural data sources, e.g. online searches (e.g. about a product, a service or any other types of information) and online page views; and
- Opinion data sources, e.g. comments on social media.

Benefits

Apart from providing valuable information, Big Data can possibly bring down the cost in data collection, as Big Data is usually generated digitally and requires no data collection by surveys or interviews. There is no "respondent burden" associated with Big Data, as it is a spontaneous data source. Also, Big Data can provide additional data variables for better stratification, weighting and estimation to complement existing data sources for compiling official statistics. They can also be used to validate estimates produced using traditional methods.

Due to the frequent-disseminating nature and high volume of Big Data, it can help provide more timely estimates of statistics, in contrast to traditional data collection methods such as surveys, which typically take a long time to collect and process the data.

Challenges / Limitations

While Big Data could be potentially useful, there are also a number of challenges that require attention. Very often, Big Data cannot be used to generalise the observations to the target population of interest, as the data sources may cover only a subset of the target population (and often not a random sample). For example, in the case of social media data, the majority of the users are younger generation. Thus the analysis carried out using such data cannot represent the whole population.

Big Data analysis requires statistical methods such as data mining and text mining, which are different from how official statistics are traditionally compiled. Adjustments to statistical results from Big Data might be needed as datasets from Big Data sources are not necessarily random samples of the target population. Moreover, underlying relationships unveiled from Big Data could be spurious correlation that occurs by chance. It will be misleading to conclude that such correlation pattern is a universal phenomenon applicable to the whole population.

As Big Data is often unstructured and in large volume, specialised data management infrastructure and tools are required. The implementation, operation and management costs of such specialised infrastructure and tools could be high. Specialists with expertise in Big Data software frameworks, programming and adequate statistical knowledge are also needed to manage the infrastructure and data warehouses, and to clean and analyse the data.

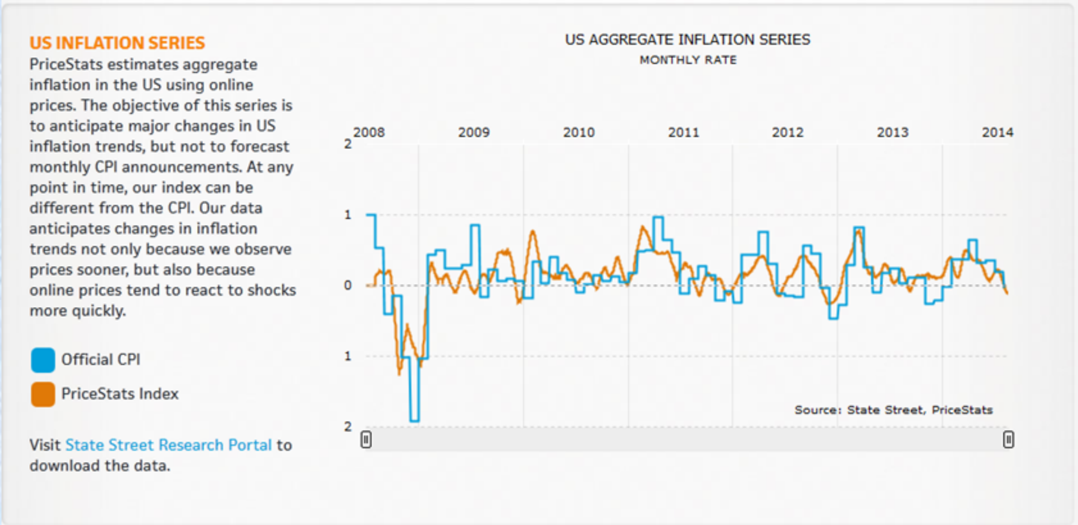
Last but not least, as Big Data often contains personal information, it is necessary to govern the use of such data in order to protect privacy of the data subjects.

Case Studies

In this section, several cases of Big Data applications related to official statistics as seen around the world will be reviewed.

1 **Price statistics**

PriceStats is an organisation that uses online price data to compile daily inflation statistics. By making use of web scraping technologies, data are automatically collected from the Internet, which are then converted from unstructured data in HTML format into structured datasets. PriceStats collects price data across selected retailers over 70 countries, and with these price data, price statistics for over 20 countries such as the United States, Australia, Japan and China are published, covering a range of economic sectors (such as recreation & culture, food & beverages, clothing & footwear etc.). Comparing the US official CPI series and PriceStats Index (Figure 1), the two have very similar patterns, suggesting that price statistics compiled using Big Data can supplement the price indices generated by traditional method. Using the Big Data approach has the advantages that the collection cost is cheaper (without manual collection) and the dissemination is more timely (as traditional CPI is released only monthly, but Big Data price statistics can be released daily).



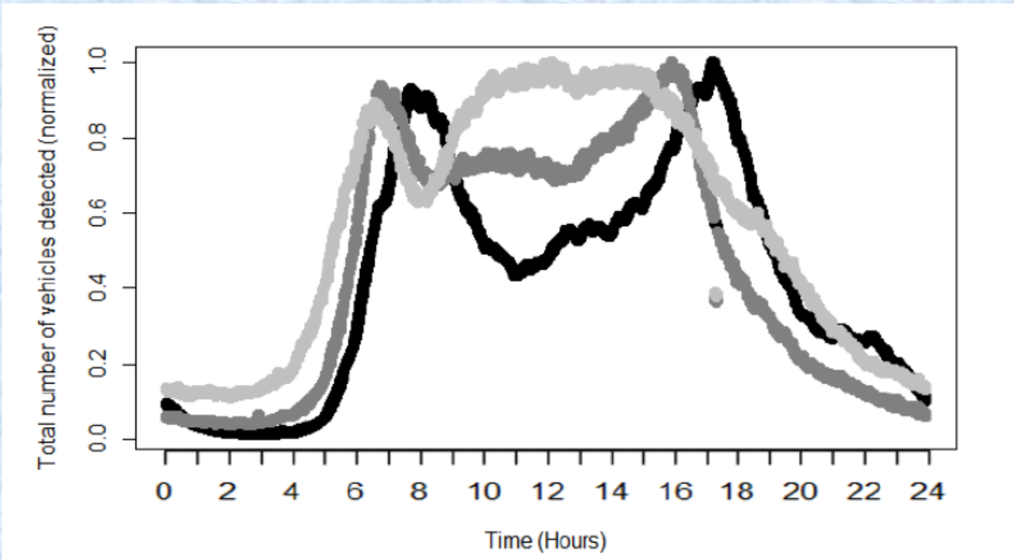
◄ *Figure 1: US Aggregate Inflation Series (Official CPI and PriceStats Index).*

2 Traffic and transport statistics

The vast amount of Big Data collected on the roads by cameras, sensors and GPS can provide valuable insights on the traffic conditions. In Ireland, the Dublin City Council (DCC) has conducted a research with IBM to integrate, process and visualise daily transportation data. Real-time geospatial data collected using GPS from 1 000 buses across Dublin and data from bus timetables are integrated into a central geographic information system. The traffic control centre can monitor the current status of the entire bus network through a real-time digital map and dashboard. Also, detailed reports on areas of the network with frequent bus delays are generated for further action to ease congestion. The DCC and IBM are currently working to integrate meteorological data into their traffic control solution, so that alerts can be sent to controllers on potential hazards and dangers on the road.

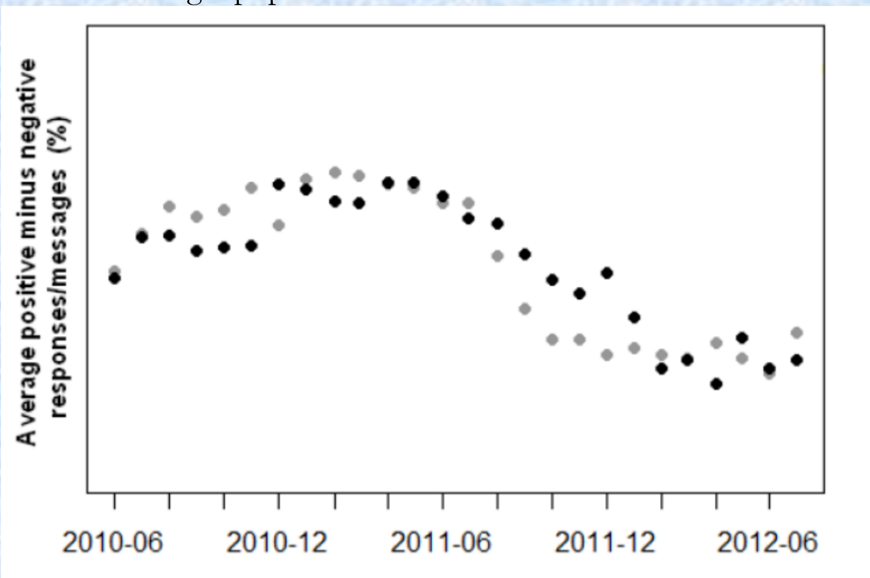
In the Netherlands, a study on traffic data was conducted by Statistics Netherlands. Some 100 million records of traffic loop detection data in a day were detected by over 10 000 traffic loop sensors, which were then harvested into a dataset containing attributes such as time, length of vehicles and location. By analysing this dataset, patterns of vehicle counts (by vehicle size) at various time points during a day could be identified (Figure 2). This can help understand the driving behaviours of vehicles of different sizes in a typical day. As noted by Statistics Netherlands, making full use of this data will result in speedier and more robust statistics on traffic.

► *Figure 2: Normalised number of vehicles detected in three length categories on December 1st, 2011 after correcting for missing data. Small (<= 5.6 meter) in black, medium sized (>5.6 and <= 12.2 meter) in dark grey, and large vehicles (> 12.2 meter) in grey.*



3 Social media

The proliferation of social media generates a lot of data from activities of social media users, which in turn provides a valuable source of Big Data. Statistics Netherlands had conducted a research project to study the sentiment of messages on social media sites including Twitter, Facebook and Google+ etc. The team obtained access to social media messages through infrastructure provided by a commercial company and used general terms to search messages with sentiment. The sentiment of each of the 30 million messages in each month was automatically classified into three groups (viz. positive, negative or neutral) as determined by the number of positive and negative words. Using this dataset, a monthly sentiment series was compiled, which was found to be highly correlated with the official series of Dutch consumer confidence and the sentiment indicator towards the economic climate (Figure 3). As noted by Statistics Netherlands, social media data has the shortcoming of representativeness, as not everybody uses social media. If background information of social media users is available, it can be used to assess representativeness through comparison with that of the target population.



► *Figure 3: Dutch consumer confidence (in grey) and the overall sentiment in Dutch social media messages (black) on a monthly basis.*

Conclusion

Big Data is a valuable source of supplementary information to complement existing official statistics that are compiled mostly through surveys. To make full use of Big Data, challenges, such as coverage, methodology, how to obtain the data, data privacy and new technical skills and infrastructure required to manage the data will have to be addressed.

Reference

The United Nations Economic and Social Council (2014). *"Big data and modernization of statistical systems: Report of the Secretary-General".*

Capps, C., & Wright T. (2013, August 1). *"Toward a Vision: Official Statistics and Big Data", Amstat News.*

Daas, P.and Puts, M. (2014), "Big data as a source of statistical information", *The survey Statistician* 69, 22-31

Daas, P.J.H., Puts, M.J., Buelens, B. & van den Hurk, P.A.M. (2013). *"Big Data and Official Statistics".*

Executive Office of the President, the White House. (2014, May). *"Big Data: Seizing Opportunities, Preserving Values".*

Harford, T. (2014, March 28). *"Big data: are we making a big mistake?", Financial Times.*

IBM Ireland (2013 July). *"Dublin City Council: Leveraging the leading edge of IBM Smarter Cities".*

Itou, T. (2014). *"Big Data and Official Statistics: Analysis of Recent Discussions in Statistical Communities".*

Savitz, E. (2012 April 16). *"The Big Cost Of Big Data", Forbes.*

Scannapieco, M., Virgillito, A. & Zardetto, D. (2012). *"Placing Big Data in Official Statistics: A Big Challenge?".*

The 2014/15 Statistical Project Competition for Secondary School Students

Mr. Alan CHEUNG
Chairperson, Organising Committee of the Competition

The Statistical Project Competition (SPC) for Secondary School Students has been an annual event of the Hong Kong Statistical Society (HKSS) since 1986/87. The 2014/15 SPC is jointly organised by the Hong Kong Statistical Society and the Education Bureau and sponsored by the Hang Seng Indexes Company Limited.

The 2014/15 SPC is the 29th round. The three main objectives of the Competition are: (i) To promote the interest of secondary school students in research methods, statistics and statistical techniques; (ii) To encourage students to understand the local community in a scientific and objective manner through the proper use of statistics; and (iii) To promote a sense of civic awareness among the students.

The Competition is divided into two Sections, namely Junior Section for Secondary 1 to 3 students and Senior Section for Secondary 4 to 6 students. Junior Section participants must submit a poster, while Senior Section participants must submit a project in the form of a written report. In addition to the First, Second, Third and Distinguished Prizes, the Competition also has the Hang Seng Indexes Company Limited Prizes for the Best Index Application and the Department of Management Sciences, The City University of Hong Kong Prizes for the Best Graphical Presentation of Statistics.

A briefing seminar for the 2014/15 SPC was held on 25 October 2014 to give an introduction of the Competition for interested schools and students. Our Chief Adjudicator, Professor WONG Heung of The Hong Kong Polytechnic University, briefed the participants and supervising teachers on the adjudication process and criteria. Ms. Carly LAI, the representative from the Census and Statistics Department, gave a brief account on Data Sources and Data Analysis. In addition, the winning teams of the 2013/14 SPC shared their experiences in the briefing seminar.

In this round, the SPC has received enthusiastic responses, with the participation of 131 statistical projects from 531 students of 45 secondary schools. The projects cover a wide variety of themes, focusing on various social and economic aspects of Hong Kong. Contemporary issues studied by participating students include demographic trends, environmental protection issues and Hong Kong's economic development, etc.

Regarding the results of the Competition, students of Shun Tak Fraternal Association Lee Shau Kee College, who used statistics to objectively analyse why the divorced female population was higher than the divorced male population, won the First Prize of the Junior Section. Students of Stewards Pooi Kei College won the Second Prize as well as the Prize for the Best Graphical Presentation of Statistics, while students of Po Leung Kuk Tong Nai Kan Junior Secondary College won the Third Prize. Students of The Chinese Foundation Secondary School won the Prize for the Best Index Application.

As for the Senior Section, the statistical report from students of Good Hope School was appraised as the best amongst all the projects. They applied official statistics from multiple facets to study the prospects of youngsters in Hong Kong, on such aspects as education prospect, career opportunities and property ownership. Students of Hong Kong Sze Yap Commercial and Industrial Association Wong Tai Shan Memorial College and Shun Lee Catholic Secondary School won the Second Prize and Third Prize respectively. Students of Sha Tin Methodist College won the Prize for the Best Graphical Presentation of Statistics while students of Chiu Lut Sau Memorial Secondary School won the Prize for the Best Index Application.

A prize presentation ceremony was held on 27 June 2015 at the West Kowloon Campus, The Hong Kong Polytechnic University. Officiating guests included Dr. Catherine CHAN, Deputy Secretary for Education and Mr. Leslie TANG, Commissioner for Census and Statistics. During the Ceremony, Professor NG Kai Wang, President of the HKSS, gave the opening address. Mr. Daniel WONG, Director, Head of Research and Development, Hang Seng Indexes Company Limited, the sponsor of the Competition, presented the prizes and Miss Viviane HO, Senior Vice President of Research and Development gave a talk on Hang Seng Index and Statistics to participants. Dr. Geoffrey TSO, Associate Head, Department of Management Sciences, The City University of Hong Kong also presented the prizes.

▲ **Photo of the officiating guests of the 2014/15 SPC with some honourable guests (from left): Miss Viviane HO, Dr. Geoffrey TSO, Mr. Daniel WONG, Dr. Catherine CHAN, Mr. Leslie TANG, Professor NG Kai Wang NG and Mr. Alan CHEUNG.**

▲ **Photo of the officiating guests with the Organising Committee of the 2014/15 SPC.**

I would like to take this opportunity to express my gratitude to all the members of the Organising Committee and the adjudicators of the 2014/15 SPC for their support and assistance. Last but not least, I would also like to thank the patrons of the Competition, Dr. Catherine CHAN, Deputy Sectary for Education and Mr. Leslie TANG, Commissioner for Census and Statistics; the sponsor of the Competition, the Hang Seng Indexes Company Limited; as well as the Department of Management Sciences of The City University of Hong Kong which had also rendered financial support to the event.

Report on the
Statistics Creative-Writing Competition for 2014/15

Dr. Philip L.H. YU

Chairperson, Organising Committee of the Competition

To raise the interest of young students in statistics and its application, the Hong Kong Statistical Society and Education Bureau jointly organised the Statistics Creative-Writing Competition (SCC) for Secondary School Students for the first time in 2009. The key objectives of the Competition are to raise the interest of students in statistics and its application; and to encourage students to creatively express in words the daily application of statistical concepts or incorporate statistical concepts into a story in a scientific and objective manner.

The 2014/15 SCC is the 6th round of the Competition. A briefing seminar was held on 25 October 2014 to give an introduction of the Competition for interested schools and students. The supervising teacher of the champion team in 2013/14 SCC was invited to share his experience in guiding his team in the Competition. Also, Dr. CHUNG Yuk-ka of The University of Hong Kong was invited to give an introduction on the theme topic "Linear regression".

The 6th round of the SCC was completed with a total of 45 entries received. Altogether, 98 students from 30 schools participated in the Competition. An adjudication panel, which was led by Dr. CHEUNG Ka-chun of The University of Hong Kong and comprised colleagues from mathematical education section of the Education Bureau, university teaching staff and professional statisticians, was set up for the Competition. Having undergone stringent scrutiny by the adjudicators, outstanding entries were selected for receiving the awards finally.

Dr. CHUNG Yuk-ka introduced the theme topic "Linear regression" in the briefing session of the 2014/15 SCC.

The prize presentation ceremony was held on 27 June 2015 at the West Kowloon Campus, The Hong Kong Polytechnic University (PolyU). The guests invited to officiate at the ceremony are Mr. Leslie TANG, Commissioner of the Census and Statistics Department, Ms. CHING Suk-yee, Principal Education Officer (Curriculum Development) of the Education Bureau and Dr. Samuel CHAN, Division Head of Business of the Hong Kong Community College.

▲ **Photo of the winning teams and the officiating guests in the 2014/15 SCC Prize Presentation Ceremony.**

Apart from the champion, runner-ups, distinguished prizes and entry prizes, a thematic prize, named as "Department of Statistics & Actuarial Science, The University of Hong Kong Prize for the Best Thematic Writing", has been introduced since the 2011/12 round. The theme topics set so far are "Correlation", "Outlier", "Missing Data" and "Linear Regression". In addition to this, the PolyU Hong Kong Community College has sponsored a new thematic prize in the 2014/15 SCC, namely the "PolyU Hong Kong Community College Prize for the Best Article Presentation".

In line with the development of mathematics education in Hong Kong, and providing teachers with more references, booklets of winning entries in the SCC had been published. To enhance the overall quality of the booklets, the booklets also included several invited articles written by university professors, school teachers and statisticians from the Census and Statistics Department. We believe that through reading a series of interesting articles and creative stories in the booklet, many students will be able to gain knowledge of statistics and recognised how important the proper use of statistical concepts in analysing problems. The booklets are issued for free distribution to secondary schools and are available in the HKSS website for free download.

▲ **Photo of the officiating guests and the Organising Committee in the 2014/15 SCC:**
(Front row from left) Dr. Samuel CHAN; Mr. FUNG Hing-wang; Ms. CHING Suk-yee; Professor NG Kai Wang NG; Mr. Leslie TANG; Dr. Wilson KWAN and Mr. P.L. LEE.

I would like to take this opportunity to express my gratitude to all the members of the Organising Committee and the adjudicators in the 2014/15 SCC for their help and support. Their strenuous efforts have undoubtedly contributed to enhancing students' statistical literacy and raising their interest in statistics. I would also like to thank the Department of Statistics and Actuarial Science of The University of Hong Kong and the PolyU Hong Kong Community College for sponsoring the prizes in the Competition.

NEWS

New Appointment

Professor JIANG Bin-yan has joined the Department of Applied Mathematics of The Polytechnic University of Hong Kong with effect from 1 August 2015.

Promotion

- Professor POON Wai-yin of the Department of Statistics of The Chinese University of Hong Kong (CUHK) was appointed to Pro-Vice-Chancellor/ Vice-President of CUHK with effect from 1 May 2015.
- Professor SONG Xin-yuan of the Department of Statistics of CUHK was advanced to Professor with effect from 1 August 2015.
- Professor WONG May Chun-mei of the Department of Dentistry of The University of Hong Kong was advanced to Professor with effect from 1 August 2015.

Award

- Professor ZHAO Xing-qiu from the Department of Applied Mathematics of The Polytechnic University of Hong Kong was awarded the 2014 Natural Science Award (Second Class) of the Higher Education Outstanding Scientific Research Output Award by The Ministry of Education of China. His project was entitled "Study on Large Deviation Theory in Stochastic Process and Statistical Inference".
- Professor SHAO Qi-man of CUHK and Professor JING Bing-yi of The Hong Kong University of Science and Technology were jointly awarded the 2015 State Natural Science Award (Second Class) by The State Council of China. The project concerned was entitled "Self-normalised Limit Theory and Stein's Method".

NEWS

◆ World Statistics Day 2015

The United Nations General Assembly designated 20 October 2015 as the second World Statistics Day to acknowledge the importance of high quality official statistics to making of better decisions and improving the lives of people. The slogan of the World Statistics Day 2015 is "Better data. Better lives", which conveys the idea that the ultimate goal of producing high quality official statistics is to improve the lives of people. In addition, the General Assembly decided to celebrate the World Statistics Day on 20 October every five years.

The Hong Kong Statistical Society (HKSS) jointly organised a public lecture on "Better Data for Better Lives" with the Census and Statistics Department and the Education Bureau in celebration of the second World Statistics Day at the Lecture Hall of Hong Kong Science Museum on 20 October 2015. The public lecture received favourable responses from the participants.

▲ **The public lecture was well-received by the participants.**

◆ World Statistics Day 2015 (Cont'd)

Professor NG Kai Wang, President of the HKSS, was one of the honourable officiating guests who gave an opening remark at the public lecture. Professor CHAN Ngai-hang represented the HKSS in giving a talk on the topic "Statistical Finance: Everything you want to know but are afraid to ask" in the public lecture.

▲ In the public lecture celebrating the second World Statistics Day on 20 October 2015, Professor NG welcomed the guests and audience.

▲ Professor CHAN represented the HKSS in giving one of the three talks in the public lecture.

▲ Group photo of the guests of honour and speakers (from left):

Ms. Marion CHAN, Ms. Reddy NG, Dr. Catherine CHAN, Professor CHAN Ngai-hang, Professor NG Kai Wang and Mr. Leslie TANG.

◆ The HKSS has signed up to a Statement for World Statistics Day

The Royal Statistical Society (RSS), the American Statistical Association and the International Statistical Institute launched a Statement for the World Statistics Day to recognise the importance of statistics in the international development process. The HKSS has signed up to this Statement. As one of the signatories to the Statement, the HKSS supports the call for a data revolution and recognises the importance of data for policy making and for accountability in all countries of the world.

More details of the Statement can be found at the following website:

"http://www.rss.org.uk/RSS/Influencing_Change/World_Statistics_Day_statement/RSS/Influencing_Change/World_Statistics_Day_statement.aspx?hkey=c261be5f-4079-4392-aa0e-35f6021aee6c"

◆ Withdrawal from offering the HKSS exam after the May 2017 round

The RSS will discontinue its professional examinations from September 2017, after a strategy review to take forward the work of RSS in the next five years. Instead of organising professional examinations, RSS will move to an accreditation model after 2017.

Following the decision of RSS, the HKSS will withdraw from offering its professional examinations after the 2017 round of the examination. While it is understood that RSS is still deliberating the detailed accreditation model, the HKSS will follow up with RSS with a view to exploring the feasibility of setting up a similar accreditation model in Hong Kong. We will keep members posted of the latest development in good time.

NEWS

THE HKSS'S PARTICIPATION IN UPCOMING CONFERENCES

◆ *The 10th ICSA international conference*

The HKSS will co-sponsor the 10th ICSA (International Chinese Statistical Association) international conference to be held at Xuhui campus of Shanghai Jiao Tong University in China during 19-22 December 2016. Thanks to the organising committee, the HKSS is also invited to organise up to two invited sessions in the conference programme as a sponsor without financial commitment.

Besides, the HKSS President to be elected for the 2016-2017 session in the 2016 AGM has been invited to join as a panelist (who may also give a talk in another session in addition to this one) in the following special panel session on "Global Statistical Collaborations: Opportunities, Challenges, and Future" and the President for 2016-2017 may nominate his/her representative, if not being able to attend the aforesaid special session.

◆ *The symposium on "Frontier Research in Statistics and Data Science"*

The HKSS has been granted the honour to be a sponsor, without financial commitment, of the symposium titled "Frontier Research in Statistics and Data Science" to be organised by the Department of Applied Mathematics of The Hong Kong Polytechnic University during 25-26 June 2016. The symposium will cover a wide range of emerging research topics in Statistics, and provide a wonderful opportunity for statisticians to exchange ideas. The HKSS is invited to organise an invited session in the symposium.

◆ *The Fourth IMS Asia Pacific Rim Meeting*

The Fourth IMS Asia Pacific Rim Meeting (IMS-APRM), which is a meeting series of the Institute of Mathematical Statistics (IMS), will take place in Hong Kong during 27-30 June 2016. It will provide an excellent forum for scientific communications and collaborations for the researchers in Asia and Pacific Rim, and promote communications and collaborations between the researchers in this area and those from other parts of the world. Thanks to the organising committee, the HKSS will be an institutional sponsor of the IMS-APRM 2016 without financial commitment. More details of the meeting can be found at http://ims-aprm2016.sta.cuhk.edu.hk/.

◆ Seminar on "Application of Big Data Technologies in Business Industries"

The HKSS organised a public seminar on Big Data titled "Application of Big Data Technologies in Business Industries" on 24 February 2016 at The University of Hong Kong. Three speakers from SAS Institute Ltd, Mr. Kenneth KOH, Senior Manager of Regional Customer Intelligence; Mr. Jason LOH, Product Manager of Regional Information Management; and Mr. Daniel SHUM, Senior Manager of Analytic Platform, SAS HK, shared with us the applications of big data technologies in Insurance, Telco & Retails industry on customer engagement; social media & unstructured data analysis; and edge/real time analytic on performance monitoring of electric car, etc. About 35 participants attended the seminar.

▼► Some snapshots of the speakers and audience taken during the seminar.

◆ *One-day Tour on 13 March 2016*

A one-day tour was organised by the HKSS on 13 March 2016. A total of 26 members and their families/friends participated. All participants had a good time gathering and seeing the beautiful scenery of Grass Island, Hoi Ha Wan Marine Park and Shui Long Wo. Some snapshots of the day tour are given below to highlight this enjoyable social activity.