ทศพร แถลงธรรม : การจำลองรูปแบบความสูญเสียสำหรับการประกันวินาศภัย ด้วยรูปแบบผสมจำกัดของข้อมูลรายเดี่ยว (THE MODELING OF LOSS FOR NON-LIFE INSURANCE WITH FINITE MIXTURE MODELS OF INDIVIDUAL DATA)
อาจารย์ที่ปรึกษา : ศาสตราจารย์ ดร.ไพโรจน์ สัตยธรรม, 142 หน้า.

การวิจัยครั้งนี้มีวัตถุประสงค์เพื่อศึกษาหารูปแบบของความสูญเสียทางด้านประกันวินาศภัย สำหรับข้อมูลรายเดี่ยวที่มีรูปแบบเป็นแบบผสม และใช้รูปแบบที่เหมาะสมนั้นไปกำหนดเบี้ยประกันภัย สามารถสรุปผลการศึกษาได้ดังต่อไปนี้

การศึกษารูปแบบของความสูญเสีย (ค่าสินไหมทดแทน) ทางด้านประกันวินาศภัย แบ่งออกได้เป็น 2 ส่วน ตามรายการที่แสดงข้างล่างดังนี้

ส่วนที่ 1: การจำลอง: สำหรับรูปแบบของการแจกแจงแบบเดี่ยวของลอกนอร์มอล ใช้วิธีการประมาณค่าพารามิเตอร์ คือ วิธีภาวะน่าจะเป็นสูงสุด (maximum likelihood estimate : MLE) ข้อมูลเชิงการทดลองมี 3 กลุ่มด้วยกัน คือ ข้อมูลที่เกิดจากการจำลองการผสมของส่วนประกอบ (components) ข้อมูลที่มีการแจกแจงของความสูญเสีย (empirical data which are simulated by mixed components of loss distributions : EMD) การผสมของส่วนประกอบข้อมูลการแจกแจงความสูญเสียแบบคอมพาวด์ปัวส์ซองที่มีอัตราส่วนลดของอัตราดอกเบี้ย (mixed components of discounted compound Poisson-mixed loss distributions with interest rate : EDP) และข้อมูล EMD ด้วยเทคนิค bootstrap สำหรับรูปแบบของการแจกแจงแบบผสมจำกัดของลอกนอร์มอล ใช้วิธีการประมาณค่าพารามิเตอร์ คือ Expectations Maximization (EM) algorithm และใช้ข้อมูล EMD เป็นข้อมูลเชิงการทดลอง

การทดสอบภาวะสารูปสนิทดี (GOF) ที่ใช้วัดการเทียบเคียงกันได้ของกลุ่มตัวอย่างสุ่มกับฟังก์ชันการแจกแจงทางทฤษฎีนั้น เป็นวิธี Kolmogorov-Smirnov test (*K-S* test) และ Anderson-Darling test (*A-D* test)

การแจกแจงความสูญเสีย ประกอบด้วย การแจกแจงลอกนอร์มอล แกมม่า พาเรโต และไวบูลล์ ข้อมูลที่ใช้ในการทดลองนี้ จำลองโดย MATLAB ซึ่งกระทำซ้ำกัน 200 ครั้ง ในแต่ละกรณี

ผลการศึกษาการจำลอง: สำหรับทุก ๆ ขนาดตัวอย่าง พบว่า ข้อมูล EMD ข้อมูล EDP และข้อมูล EMD ด้วยเทคนิค bootstrap ไม่สามารถมีความสอดคล้องเหมาะสมกับการแจกแจงลอกนอร์มอล สำหรับรูปแบบการแจกแจงแบบผสมจำกัดของลอกนอร์มอลนั้น สามารถมีลักษณะสอดคล้องเหมาะสมกับข้อมูล EMD ที่ทำการจำลองทุกกรณี ด้วยระดับนัยสำคัญที่ 0.10 การยอมรับของลักษณะสอดคล้องเหมาะสมนี้จะมีมากขึ้นตามจำนวนของส่วนประกอบ (k) ที่เพิ่มขึ้นด้วย

ส่วนที่ 2: พิจารณาข้อมูลการจ่ายค่าสินไหมทดแทนของการประกันภัยรถยนต์ข้อมูลรายเดี่ยวในปี 2552 ของบริษัทประกันวินาศภัยแห่งหนึ่งในประเทศไทย ผลการศึกษาพบว่าการประกันภัยประเภทความคุ้มครองที่ 5 จำนวน 1,296 ข้อมูล มีลักษณะสอดคล้องเหมาะสมกับการแจกแจงแบบผสมจำกัดของลอกนอร์มอล ด้วยการทดสอบ *K-S* และ *A-D* มีระดับนัยสำคัญที่ 0.10 โดยจำนวนส่วนประกอบ (k) ที่เพิ่มขึ้นจะทำให้การยอมรับลักษณะสอดคล้องนี้มากยิ่งขึ้น

การกำหนดเบี้ยประกันภัย: ได้มีการเสนอหลักการคำนวณเบี้ยประกันภัยแบบลอกทรานส์ฟอร์ม (Log-transform) ที่เกี่ยวข้องกับรูปแบบการแจกแจงแบบผสมจำกัดของลอกนอร์มอล ซึ่งหลักการคำนวณนี้จะช่วยแก้ไขปัญหาในการบริหารการจัดการที่เกิดขึ้นจริงในทางปฏิบัติ เมื่อนำหลักการคำนวณเบี้ยประกันภัยแบบลอกทรานส์ฟอร์ม มาประยุกต์ใช้กับการประกันภัยรถยนต์ประเภทความคุ้มครองที่ 5 ผลการศึกษาพบว่า เบี้ยประกันภัยที่คำนวณด้วยวิธีแบบลอกทรานส์ฟอร์ม จะให้ค่าเบี้ยประกันภัยที่ต่ำกว่าเบี้ยประกันภัยที่มีการคำนวณตามวิธีการอื่น ๆ เช่น เบี้ยประกันภัยสุทธิ (net) เบี้ยประกันภัยค่าคาดหวัง (expected value) เบี้ยประกันภัยเบี่ยงเบนมาตรฐาน (standard deviation) และเบี้ยประกันภัยแบบหวางทรานส์ฟอร์ม (Wang transform) เบี้ยประกันภัยที่มีค่าน้อยที่สุดตามวิธีแบบลอกทรานส์ฟอร์ม คือเบี้ยประกันภัยที่คำนวณด้วย $k = 100$ ซึ่งการคำนวณตามวิธีแบบลอกทรานส์ฟอร์มนี้ จะเป็นประโยชน์สำหรับหลักการตัดสินใจของบริษัทในการกำหนดเบี้ยประกันภัยได้เป็นอย่างดี

สาขาวิชาคณิตศาสตร์ ลายมือชื่อนักศึกษา ______________________

ปีการศึกษา 2555 ลายมือชื่ออาจารย์ที่ปรึกษา ________________

TOSAPORN TALANGTAM : THE MODELING OF LOSS FOR NON-LIFE INSURANCE WITH FINITE MIXTURE MODELS OF INDIVIDUAL DATA. THESIS ADVISOR : PROF. PAIROTE SATTAYATHAM, Ph.D. 142 PP.

BOOTSTRAP / CLAIM SIZE DISTRIBUTION / EM ALGORITHM / EQUILIBRIUM PRICE / FINITE MIXTURE MODELS / LOG-TRANSFORM / LOGNORMAL DISTRIBUTION / LOSS DISTRIBUTION / MOTOR INSURANCE / PREMIUM CALCULATION PRINCIPLES / WANG TRANSFORM

The objective of this study is to find loss distribution models for mixture models of individual data and use a suitable model to price the insurance premium. The results of the study are as follows:

The modeling of loss (claim) for non-life insurance: It is separated into 2 parts as shown below.

Part 1: The simulations: For the model of a single parametric Lognormal distribution, the parameter estimation is the maximum likelihood estimate (MLE). There are 3 sets of empirical data for fitting, namely, the empirical data which are simulated by mixed components of loss distributions (EMD), mixed components of discounted compound Poisson-mixed loss distributions with interest rate (EDP) and the EMD with the bootstrap technique. For the model of finite mixture Lognormal distributions, the estimated parameters of the model are obtained from Expectations Maximization (EM) algorithm and the empirical data for fitting is EMD.

The goodness of fit (GOF) test measures the compatibility of a random sample with a theoretical probability distribution function. We use the Kolmogorov-Smirnov test (*K*-*S* test) and the Anderson-Darling test (*A*-*D* test).

The loss distributions are Lognormal, Gamma, Pareto and Weibull. Data sizes are obtained through simulation using MATLAB and repeated 200 times in each case.

The simulation results: For any sample size, we found that the EMD, EDP and EMD with the bootstrap technique cannot be fitted by any Lognormal distribution. For the model of finite mixture Lognormal distributions, they can be fitted to EMD in any case of simulation with a significance level of 0.10. This fitting is better when the number of components (k) are increased.

Part 2: we consider the individual data for motor insurance claims for the year 2009 from a non-life insurance company in Thailand. We found that 1,296 observations of type - 5 meet the mixture Lognormal distributions at a significant level of 0.10 for both the *K*-*S* and *A*-*D* tests. The fitting is better when the number of components (k) are increased.

The insurance pricing: We introduce the Log-transform premium principle related to the finite mixture Lognormal distributions which can assist in the solving of these real world management problems. We applied the Log-transform premium principle to price motor insurance claims of type - 5 and found that the premiums based on Log-transform are less than the premiums based on some other principles: such as net, expected value, standard deviation and the Wang transform. The premium of $k = 100$ is the minimum. This is, therefore, a very useful method for providing a sound basis for company decisions on premium pricing.

School of Mathematics — Student's Signature__________________

Academic Year 2012 — Advisor's Signature__________________

ACKNOWLEDGEMENTS

I am grateful to Prof. Dr. Pairote Sattayatham for his lectures on mathematical finance and non-life insurance mathematics. He has given me much useful advice and valuable suggestions on the mathematics and statistics in this thesis.

I very much appreciate the assistance of Asst. Prof. Dr. Eckart Schulz for his lectures on mathematical analysis and for his careful reading, helpful comments and checking of my manuscript. I would also like to thank Assoc. Prof. Dr. Samruam Choungchareon for his many useful comments and technical suggestions. My thanks also go to Assoc. Prof. Dr. Prapasri Asawakun and Asst. Prof. Dr. Arjuna Chaiyasena for their helpful comments and giving me encouragement. I am greatly indebted to all the lecturers in the School of Mathematics for all the lectures I attended while studying.

Finally, special thanks are due to my parents: Colonel Tongpan and Pensri Talangtam, and to my friends for their encouragement and moral support throughout my study period.

Tosaporn Talangtam

CONTENTS

CONTENTS (Continued)

CONTENTS (Continued)

LIST OF TABLES

LIST OF TABLES (Continued)

LIST OF TABLES (Continued)

Table **Page**

LIST OF TABLES (Continued)

Table **Page**

LIST OF TABLES (Continued)

LIST OF FIGURES

ABBREVIATIONS AND SYMBOLS

CDF	cumulative distribution function
PDF	probability density function
MLE	maximum likelihood estimate
EM	expectations maximization algorithm
MGF	moment generating function
i.i.d.	independent, identically distributed
ω	elementary event
Ω	sample space
n	sample size (policy number)
$F(x)$	CDF
$F_n(x)$	empirical CDF
$f(x)$	PDF
X	random variable of loss (claim)
$F_X(x)$	CDF of X
$f_X(x)$	PDF of X
$E[X]$	expected value of X
$Var[X]$	variance of X
$Corr[X,Y]$	correlation coefficient between X and Y
$Cov[X,Y]$	covariance between X and Y
$\rho(X,Y)$	$Corr[X,Y]$

ABBREVIATIONS AND SYMBOLS (Continued)

$\boldsymbol{x} = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}$	a point of $\mathbb{R}^n$ is an n dimensional vector
$\boldsymbol{x}'$	a row vector with components $x_1, \ldots, x_n$.
	The symbol $'$ is used to indicate transposition.
I_A	indicator of the event A
$M_X(t)$	MGF
$N(\mu, \sigma^2)$	normal distribution with mean μ and variance σ^2
$LN(\mu, \sigma)$	Lognormal distribution with parameters μ and σ
$\Phi(x)$	standard normal distribution function
$\mathbb{R}$	the set of real numbers
$n!$	$(n)(n-1)\cdots(1)$