SOME TOPICS RELEVANT TO ACTUARIAL MATHEMATICS

by

Yu Jin

M.S., Southern Illinois University Carbondale, 2010

A Dissertation
Submitted in Partial Fulfillment of the Requirements for the
Doctor of Philosophy Degree

School of Mathematical and Statistical Sciences
in the Graduate School
Southern Illinois University Carbondale
May 2023

DISSERTATION APPROVAL

SOME TOPICS RELEVANT TO ACTUARIAL MATHEMATICS

by

Yu Jin

A Dissertation Submitted in Partial

Fulfillment of the Requirements

for the Degree of

Doctor of Philosophy

in the field of Mathematics

Approved by:

Dr. David Olive, Chair

Dr. Xiaoxin Wang Beardsley

Dr. S. Yaser Samadi

Dr. Dashun Xu

Dr. Kwangho Choiy

Graduate School
Southern Illinois University Carbondale
February 7, 2023

AN ABSTRACT OF THE DISSERTATION OF

Yu Jin, for the Doctor of Philosophy degree in Mathematics, presented on October 19, 2022, at Southern Illinois University Carbondale.

TITLE: SOME TOPICS RELEVANT TO ACTUARIAL MATHEMATICS

MAJOR PROFESSOR: Dr. David Olive

This paper gives large sample theory for some ridge-type multiple linear regression estimators, including Liu-type regression estimators, when the number of predictors is fixed, as well as prediction intervals for censored data from some parametric models, which could be practically applied in the field of actuarial science.

KEY WORDS: Ridge Regression, Liu-Type Regression Estimator, Predictive Distribution, Shorth, Life Contingencies.

ACKNOWLEDGMENTS

I would like to thank Dr. Olive for his invaluable assistance and guidance leading to the writing of this paper. My sincere thanks also goes to the other four members of my graduate committee for their patience and understanding during the past few months of effort that went into the final production and presentation of this paper. In addition, a special thanks to Dr. Mcsorley, the former Director of Graduate Studies and Graduate Advisor, for his tremendous help and unwavering academic support.

I am also grateful for my family - my parents, Ping and Pengfeng, for their unmatched love and support throughout my life. Without them, none of my studies would be possible.

TABLE OF CONTENTS

LIST OF TABLES

CHAPTER 1

INTRODUCTION

In Actuarial mathematics, Statistics has become increasingly important since large insurance companies and banks have large data bases from which information needs to be extracted. Statistical Learning methods such as shrinkage estimators have become important. Chapter 2 derives the large sample theory for ridge type estimators such as the Liu-Type Regression Estimators.

Chapter 3 gives frequentist prediction intervals based on the maximum likelihood estimator. Actuarial texts often use Bayesian predictive distributions.

Chapter 4 gives some simple proofs for some formulas that are useful for the life contingencies actuarial exams.

CHAPTER 2

LARGE SAMPLE THEORY FOR SOME RIDGE-TYPE REGRESSION ESTIMATORS

2.1 INTRODUCTION

This section reviews the multiple linear regression model, some ridge-type regression estimators, and the large sample theory for the ordinary least squares estimator. Suppose that the response variable Y_i and at least one predictor variable $x_{i,j}$ are quantitative with $x_{i,1} \equiv 1$. Let $\boldsymbol{x}_i^T = (x_{i,1}, ..., x_{i,p})$ and $\boldsymbol{\beta} = (\beta_1, ..., \beta_p)^T$ where β_1 corresponds to the intercept. Then the multiple linear regression (MLR) model is

$$Y_i = \beta_1 + x_{i,2}\beta_2 + \cdots + x_{i,p}\beta_p + e_i = \boldsymbol{x}_i^T\boldsymbol{\beta} + e_i \tag{2.1}$$

for $i = 1, ..., n$. Here n is the sample size, and assume that the random variables e_i are independent and identically distributed (iid) with mean $E(e_i) = 0$ and variance $V(e_i) = \sigma^2$. In matrix notation, these n equations become

$$\boldsymbol{Y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{e} \tag{2.2}$$

where $\boldsymbol{Y}$ is an $n \times 1$ vector of response variables, $\boldsymbol{X}$ is an $n \times p$ matrix of predictors, $\boldsymbol{\beta}$ is a $p \times 1$ vector of unknown coefficients, and $\boldsymbol{e}$ is an $n \times 1$ vector of unknown errors. The ith fitted value $\hat{Y}_i = \boldsymbol{x}_i^T\hat{\boldsymbol{\beta}}$ and the ith residual $r_i = Y_i - \hat{Y}_i$ where $\hat{\boldsymbol{\beta}}$ is any $p \times 1$ estimator of $\boldsymbol{\beta}$. Ordinary least squares (OLS) is often used for inference if n/p is large.

Liu (2003) defined the Liu-type estimator

$$\hat{\boldsymbol{\beta}}_{k,d} = (\boldsymbol{X}^T\boldsymbol{X} + k\boldsymbol{I})^{-1}(\boldsymbol{X}^T\boldsymbol{Y} - d\hat{\boldsymbol{\beta}}) = \hat{\boldsymbol{\beta}}_{R,k} - \frac{d}{n}n(\boldsymbol{X}^T\boldsymbol{X} + k\boldsymbol{I})^{-1}\hat{\boldsymbol{\beta}} \tag{2.3}$$

where $k = k_n \geq 0$, $d = d_n$ is a real number, and the Hoerl and Kennard (1970) ridge regression estimator $\hat{\boldsymbol{\beta}}_{R,k}$ corresponds to $d = 0$. The Liu (1993) estimator

$$\hat{\boldsymbol{\beta}}_c = (\boldsymbol{X}^T\boldsymbol{X} + \boldsymbol{I})^{-1}(\boldsymbol{X}^T\boldsymbol{Y} + c\hat{\boldsymbol{\beta}})$$

is another special case with $k = 1$ and $d = -c$ where $0 < c < 1$.

Kurnaz and Akay (2015) showed that several ridge-type regression estimators in the literature can be written as $\hat{\boldsymbol{\beta}}_f = (\boldsymbol{X}^T\boldsymbol{X} + k\boldsymbol{I})^{-1}(\boldsymbol{X}^T\boldsymbol{Y} + f(k)\hat{\boldsymbol{\beta}})$ where $k \geq 0$ and $f(\cdot)$ is a continuous function of k, including ridge-type estimators given by Özkale and Kaçiranlar (2007), Sakallioğlu and Kaçiranlar (2008), and Yang and Chang (2010). Note that $\hat{\boldsymbol{\beta}}_f = \hat{\boldsymbol{\beta}}_{k,d}$ with $d = -f(k)$. If $\hat{\boldsymbol{\beta}} = \hat{\boldsymbol{\beta}}_{R,k}$, then $\hat{\boldsymbol{\beta}}_f = (\boldsymbol{X}^T\boldsymbol{X} + k\boldsymbol{I})^{-1}(\boldsymbol{X}^T\boldsymbol{X} + [k + f(k)]\boldsymbol{I})\hat{\boldsymbol{\beta}}_{R,k}$.

Kibria and Lukman (2020) defined the estimator

$$\hat{\boldsymbol{\beta}}_{KL} = (\boldsymbol{X}^T\boldsymbol{X} + k\boldsymbol{I})^{-1}(\boldsymbol{X}^T\boldsymbol{X} - k\boldsymbol{I})\hat{\boldsymbol{\beta}}_{OLS}.$$

Since $(\boldsymbol{X}^T\boldsymbol{X} + k\boldsymbol{I})^{-1}(\boldsymbol{X}^T\boldsymbol{X} - k\boldsymbol{I}) = \boldsymbol{I} - 2k(\boldsymbol{X}^T\boldsymbol{X} + k\boldsymbol{I})^{-1}$,

$$\hat{\boldsymbol{\beta}}_{KL} = [\boldsymbol{I} - 2k(\boldsymbol{X}^T\boldsymbol{X} + k\boldsymbol{I})^{-1}]\hat{\boldsymbol{\beta}}_{OLS} = \hat{\boldsymbol{\beta}}_{OLS} - 2k(\boldsymbol{X}^T\boldsymbol{X} + k\boldsymbol{I})^{-1}\hat{\boldsymbol{\beta}}_{OLS}. \tag{2.4}$$

The OLS estimator $\hat{\boldsymbol{\beta}}_{OLS} = (\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{X}^T\boldsymbol{Y}$ has large sample theory given, for example, by Sen and Singer (1993, p. 280). Let the hat matrix $\boldsymbol{H} = \boldsymbol{X}(\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{X}^T$ and let the ith leverage $h_i = \boldsymbol{H}_{ii}$ be the ith diagonal element of $\boldsymbol{H}$. Consider the multiple linear regression model (2.1) where the e_i are iid with $E(e_i) = 0$ and $\mathrm{V}(e_i) = \sigma^2$. Assume that $\max_i(h_1, ..., h_n) \to 0$ in probability as $n \to \infty$ and

$$\frac{\boldsymbol{X}^T\boldsymbol{X}}{n} \to \boldsymbol{V}^{-1}$$

as $n \to \infty$. Then

$$\sqrt{n}(\hat{\boldsymbol{\beta}}_{OLS} - \boldsymbol{\beta}) \xrightarrow{D} N_p(\boldsymbol{0}, \sigma^2\,\boldsymbol{V}). \tag{2.5}$$

Note that $n(\boldsymbol{X}^T\boldsymbol{X})^{-1} \to \boldsymbol{V}$, and if $k/n \to 0$, then

$$\left(\frac{\boldsymbol{X}^T\boldsymbol{X} + k\boldsymbol{I}}{n}\right)^{-1} = n(\boldsymbol{X}^T\boldsymbol{X} + k\boldsymbol{I})^{-1} \to \boldsymbol{V}. \tag{2.6}$$

Knight and Fu (2000) derived the large sample theory for ridge regression and the Tibshirani (1996) lasso estimator with p fixed. The following section derives some large sample theory for the Liu-type estimator $\hat{\boldsymbol{\beta}}_{k,d}$ and for $\hat{\boldsymbol{\beta}}_{KL}$.

2.2 LARGE SAMPLE THEORY

The large sample theory assumes that p is fixed and that Equation (2.5) holds for the OLS estimator. Then $\hat{\boldsymbol{\beta}}_{k,d} = (\boldsymbol{X}^T\boldsymbol{X} + k\boldsymbol{I})^{-1}(\boldsymbol{X}^T\boldsymbol{Y} - d\hat{\boldsymbol{\beta}}) =$

$$(\boldsymbol{X}^T\boldsymbol{X} + k\boldsymbol{I})^{-1}(\boldsymbol{X}^T\boldsymbol{X})(\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{X}^T\boldsymbol{Y} - d(\boldsymbol{X}^T\boldsymbol{X} + k\boldsymbol{I})^{-1}\hat{\boldsymbol{\beta}} =$$

$$\boldsymbol{A}_n\hat{\boldsymbol{\beta}}_{OLS} - d(\boldsymbol{X}^T\boldsymbol{X} + k\boldsymbol{I})^{-1}\hat{\boldsymbol{\beta}}$$

where $\boldsymbol{A}_n = (\boldsymbol{X}^T\boldsymbol{X} + k\boldsymbol{I})^{-1}(\boldsymbol{X}^T\boldsymbol{X}) = \boldsymbol{B}_n = \boldsymbol{I} - k(\boldsymbol{X}^T\boldsymbol{X} + k\boldsymbol{I})^{-1}$ since $\boldsymbol{A}_n - \boldsymbol{B}_n = \boldsymbol{0}$. This identity appears in Gunst and Mason (1980, p. 332) and was used by Pelawa Watagoda and Olive (2021) to simplify ridge regression large sample theory. Thus

$$\hat{\boldsymbol{\beta}}_{k,d} = [\boldsymbol{I} - k(\boldsymbol{X}^T\boldsymbol{X} + k\boldsymbol{I})^{-1}]\hat{\boldsymbol{\beta}}_{OLS} - d(\boldsymbol{X}^T\boldsymbol{X} + k\boldsymbol{I})^{-1}\hat{\boldsymbol{\beta}} =$$

$$\hat{\boldsymbol{\beta}}_{k,d} = \hat{\boldsymbol{\beta}}_{OLS} - \frac{k}{n}n(\boldsymbol{X}^T\boldsymbol{X} + k\boldsymbol{I})^{-1}\hat{\boldsymbol{\beta}}_{OLS} - \frac{d}{n}n(\boldsymbol{X}^T\boldsymbol{X} + k\boldsymbol{I})^{-1}\hat{\boldsymbol{\beta}}. \tag{2.7}$$

Theorem 1. Assume Equations 2.5) and 2.6) hold, and that $\hat{\boldsymbol{\beta}}$ is a consistent estimator of $\boldsymbol{\beta}$. a) If $k/\sqrt{n} \to 0$ and $d/\sqrt{n} \to 0$, then $\hat{\boldsymbol{\beta}}_{k,d}$ is asymptotically equivalent to $\hat{\boldsymbol{\beta}}_{OLS}$ with

$$\sqrt{n}(\hat{\boldsymbol{\beta}}_{k,d} - \boldsymbol{\beta}) \xrightarrow{D} N_p(\boldsymbol{0}, \sigma^2\,\boldsymbol{V}).$$

b) If $k/\sqrt{n} \to \tau \geq 0$ and $d/\sqrt{n} \to \delta$, then

$$\sqrt{n}(\hat{\boldsymbol{\beta}}_{k,d} - \boldsymbol{\beta}) \xrightarrow{D} N_p(-(\tau + \delta)\boldsymbol{V}\boldsymbol{\beta}, \sigma^2\,\boldsymbol{V}).$$

c) If $k/n \to 0$ and $d/n \to 0$, then $\hat{\boldsymbol{\beta}}_{k,d}$ is a consistent estimator of $\boldsymbol{\beta}$.

Proof. a) follows from b).

b) By Equation (2.7),

$$\sqrt{n}(\hat{\boldsymbol{\beta}}_{k,d} - \boldsymbol{\beta}) = \sqrt{n}(\hat{\boldsymbol{\beta}}_{OLS} - \boldsymbol{\beta}) - \frac{k}{\sqrt{n}}\,n(\boldsymbol{X}^T\boldsymbol{X} + k\boldsymbol{I})^{-1}\hat{\boldsymbol{\beta}}_{OLS} - \frac{d}{\sqrt{n}}\,n(\boldsymbol{X}^T\boldsymbol{X} + k\boldsymbol{I})^{-1}\hat{\boldsymbol{\beta}}$$

$$\xrightarrow{D} N_p(\boldsymbol{0}, \sigma^2\,\boldsymbol{V}) - \tau\boldsymbol{V}\boldsymbol{\beta} - \delta\boldsymbol{V}\boldsymbol{\beta} \sim N_p(-(\tau + \delta)\boldsymbol{V}\boldsymbol{\beta}, \sigma^2\,\boldsymbol{V}).$$

c) By Equation (2.7), $\hat{\boldsymbol{\beta}}_{k,d} \xrightarrow{P} \boldsymbol{\beta} - 0\boldsymbol{V}\boldsymbol{\beta} - 0\boldsymbol{V}\boldsymbol{\beta} = \boldsymbol{\beta}$.

Theorem 2. Assume Equations 2.5) and 2.6) hold. a) If $k/\sqrt{n} \to 0$, then $\hat{\boldsymbol{\beta}}_{KL}$ is asymptotically equivalent to $\hat{\boldsymbol{\beta}}_{OLS}$ with

$$\sqrt{n}(\hat{\boldsymbol{\beta}}_{KL} - \boldsymbol{\beta}) \xrightarrow{D} N_p(\mathbf{0}, \sigma^2 \, \boldsymbol{V}).$$

b) If $k/\sqrt{n} \to \tau \geq 0$, then

$$\sqrt{n}(\hat{\boldsymbol{\beta}}_{KL} - \boldsymbol{\beta}) \xrightarrow{D} N_p(-2\tau \boldsymbol{V}\boldsymbol{\beta}, \sigma^2 \, \boldsymbol{V}).$$

c) If $k/n \to 0$, then $\hat{\boldsymbol{\beta}}_{KL}$ is a consistent estimator of $\boldsymbol{\beta}$.

Proof. a) follows from b).

b) By Equation (2.4),

$$\sqrt{n}(\hat{\boldsymbol{\beta}}_{KL} - \boldsymbol{\beta}) = \sqrt{n}(\hat{\boldsymbol{\beta}}_{OLS} - \boldsymbol{\beta}) - \frac{2k}{\sqrt{n}} \, n(\boldsymbol{X}^T\boldsymbol{X} + k\boldsymbol{I})^{-1}\hat{\boldsymbol{\beta}}_{OLS}$$

$$\xrightarrow{D} N_p(\mathbf{0}, \sigma^2 \, \boldsymbol{V}) - 2\tau \boldsymbol{V}\boldsymbol{\beta} \sim N_p(-2\tau \boldsymbol{V}\boldsymbol{\beta}, \sigma^2 \, \boldsymbol{V}).$$

c) By Equation (2.4),

$$\hat{\boldsymbol{\beta}}_{KL} = \hat{\boldsymbol{\beta}}_{OLS} - \frac{2k}{n} \, n(\boldsymbol{X}^T\boldsymbol{X} + k\boldsymbol{I})^{-1}\hat{\boldsymbol{\beta}}_{OLS} \xrightarrow{P} \boldsymbol{\beta} - 2(0)\boldsymbol{V}\boldsymbol{\beta} = \boldsymbol{\beta}.$$

2.3 CONCLUSIONS

Theorems 1 and 2 gave some large sample theory for many ridge-type estimators. Taking $d = -k$ is interesting in Theorem 1. Several of the ridge-type estimators can be computed if $k > 0$ even if $\boldsymbol{X}^T\boldsymbol{X}$ is singular, and such estimators can be useful if $p > n$. Li and Yang (2012) gave a Liu-type estimator that replaced $\hat{\boldsymbol{\beta}}$ by a vector $\boldsymbol{b}$ that represents prior information.

For many regression estimators, a method is needed so that everyone who uses the same units of measurement for the predictors and Y gets the same $(\hat{\boldsymbol{Y}}, \hat{\boldsymbol{\beta}})$. Let the nontrivial predictors $\boldsymbol{u}_i^T = (x_{i,2}, ..., x_{i,p})$ where $\boldsymbol{x}_i = (1, \boldsymbol{u}_i^T)^T$. A common method is to use the centered response $\boldsymbol{Z} = \boldsymbol{Y} - \overline{\boldsymbol{Y}}$ where $\overline{\boldsymbol{Y}} = \overline{Y}\mathbf{1}$, and the $n \times (p-1)$ matrix of standardized nontrivial predictors $\boldsymbol{W} = (W_{ij})$. For $j = 1, ..., p-1$, let W_{ij} denote the $(j+1)$th variable standardized so that $\sum_{i=1}^n W_{ij} = 0$ and $\sum_{i=1}^n W_{ij}^2 = n$. Note that the sample correlation matrix of the nontrivial predictors $\boldsymbol{u}_i$ is $\boldsymbol{R_u} = \boldsymbol{W}^T\boldsymbol{W}/n$. Then regression through the origin is used for the model $\boldsymbol{Z} = \boldsymbol{W}\boldsymbol{\eta} + \boldsymbol{e}$ where the vector of fitted values $\hat{\boldsymbol{Y}} = \overline{\boldsymbol{Y}} + \hat{\boldsymbol{Z}}$. Large sample theory could be given for $\boldsymbol{Z} = \boldsymbol{W}\boldsymbol{\eta} + \boldsymbol{e}$, as in Pelawa Watagoda and Olive (2021), or for $\boldsymbol{Y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{e}$, as in this chapter.

[margin=1.0in]geometry

CHAPTER 3

PREDICTION INTERVALS FOR CENSORED DATA FROM SOME PARAMETRIC MODELS

3.1 INTRODUCTION

Assume that training data $Y_1, ..., Y_n$ are independent and identically distributed from a parametric distribution $Y \sim D(\boldsymbol{\theta})$ where $\boldsymbol{\theta}$ is a $d \times 1$ vector of parameters. This chapter presents a simple large sample $100(1-\delta)\%$ prediction interval (PI) for a future value Y_f given $Y_1, ..., Y_n$. Apply the nonparametric shorth prediction interval to $Y_1^*, ..., Y_B^*$ where the Y_i^* are independent and identically distributed (iid) from the distribution $D(\hat{\boldsymbol{\theta}})$. If $\hat{\boldsymbol{\theta}}$ is a consistent estimator of $\boldsymbol{\theta}$, then this prediction interval is a large sample $100(1-\delta)\%$ PI that is a consistent estimator of the shortest population interval $[L, U]$ that contains at least $1-\delta$ of the mass as $B, n \to \infty$. This PI can be regarded as a special case of the Olive, Rathnayake, and Haile (2021) prediction interval for a parametric regression model that has no predictors. Consistent estimators of $\boldsymbol{\theta}$, such as method of moments estimators, maximum likelihood estimators (MLEs) or percentile matching estimators should be used.

A large sample $100(1-\delta)\%$ prediction interval (PI) for Y_f has the form $[\hat{L}_n, \hat{U}_n]$ where $P(\hat{L}_n \le Y_f \le \hat{U}_n) \to 1-\delta$ as the sample size $n \to \infty$. A PI is asymptotically optimal if $[\hat{L}_n, \hat{U}_n] \to [L_s, U_s]$ as $n \to \infty$ where $[L_s, U_s]$ is the population shorth: the shortest interval covering at least $100(1-\delta)\%$ of the mass. (A highest density region is a union of intervals such that the sum of the lengths is minimized given at least $100(1-\delta)\%$ coverage. For a unimodal error distribution with pdf, the population shorth is the population highest density region.)

The shorth(c) estimator of the population shorth is useful for making asymptotically optimal prediction intervals if the data are iid. Let $Z_{(1)}, ..., Z_{(n)}$ be the order statistics of $Z_1, ..., Z_n$. Then let the shortest closed interval containing at least c of the Z_i be

$$\text{shorth(c)} = [\text{Z}_{(\text{s})}, \text{Z}_{(\text{s+c}-1)}]. \tag{3.1}$$

Let $\lceil x \rceil$ be the smallest integer $\geq x$, e.g., $\lceil 7.7 \rceil = 8$. Let

$$k_n = \lceil n(1-\delta) \rceil. \tag{3.2}$$

Frey (2013) showed that for large $n\delta$ and iid data, the shorth(k_n) PI has maximum undercoverage $\approx 1.12\sqrt{\delta/n}$, and used the shorth($c$) estimator as the large sample $100(1-\delta)\%$ PI where

$$c = \min(n, \lceil n[1-\delta+1.12\sqrt{\delta/n}\,]\,\rceil).$$

The large sample $100(1-\delta)\%$ PI using $Y_1^*, ..., Y_B^*$ uses the shorth(c) PI with

$$c = \min(B, \lceil B[1-\delta+1.12\sqrt{\delta/B}\,]\,\rceil). \tag{3.3}$$

The prediction interval (3.3) can have undercoverage if n is small compared to the number of estimated parameters d. The modified shorth PI (3.4) inflates PI (3.3) to compensate for parameter estimation. We want $n \geq 10d$, and the prediction interval length will be increased (penalized) if n/d is not large. Let $q_n = \min(1-\delta+0.05, 1-\delta+d/n)$ for $\delta > 0.1$ and

$$q_n = \min(1-\delta/2, 1-\delta+10\delta d/n), \quad \text{otherwise}.$$

If $1-\delta < 0.999$ and $q_n < 1-\delta+0.001$, set $q_n = 1-\delta$. Then compute the shorth PI with

$$c_{mod} = \min(B, \lceil B[q_n+1.12\sqrt{\delta/B}\,]\,\rceil). \tag{3.4}$$

Olive (2007, 2013) and Pelawa Watagoda and Olive (2021) used similar correction factors since the maximum simulated undercoverage was about 0.05 when $n = 20d$.

The prediction intervals (3.3) and (3.4) are computed using the parametric bootstrap. There are not many references for prediction intervals for parametric models with censoring. The prediction intervals tend to be constructed using predictive distributions, have complicated correction factors, lack software, and may only be applicable when $n \geq 10d$. See Hall, Peng, and Tajvidi (1999), Hall and Rieck (2001), Lawless and Fredette (2005), and Ueki and Fueda (2007).

Bayesian predictive distributions are often hard to compute. In the simplest setting, let $Z = Y_f$, let $\boldsymbol{y} = (Y_1, ..., Y_n)^T$, and let $Y_1, ..., Y_n, Z$ be iid with pdf $f_{Z|\boldsymbol{\Theta}}(z|\boldsymbol{\theta}) = f(z|\boldsymbol{\theta})$ when $\boldsymbol{\Theta} = \boldsymbol{\theta}$. Then the

Bayesian predictive pdf $f_{Z|\boldsymbol{y}}(z|\boldsymbol{y}) = \int f_{Z|\boldsymbol{\theta}}(z|\boldsymbol{\theta})\pi_{\boldsymbol{\Theta}|\boldsymbol{y}}(\boldsymbol{\theta}|\boldsymbol{y})d\boldsymbol{\theta}$ where $\pi_{\boldsymbol{\Theta}|\boldsymbol{y}}(\boldsymbol{\theta}|\boldsymbol{y})$ is the posterior pdf and $\boldsymbol{\Theta}$ is a $g \times 1$ random vector. Hence the Bayesian predictive pdf is a continuous mixing distribution weighted by the posterior pdf. See, for example, Geisser (1993), Klugman, Panjer, and Willmot (2008, p. 406), and Kellison and London (2011, pp. 409-410). Frequentists could use the pdf $f(z|\hat{\boldsymbol{\theta}}_n)$ where $\hat{\boldsymbol{\theta}}_n$ is the MLE of $\boldsymbol{\theta}$. The Bernstein-von Mises theorem, also known as the Bayesian central limit theorem, states that the posterior distribution $\boldsymbol{\Theta}|\boldsymbol{y} \approx N_g(\hat{\boldsymbol{\theta}}_n, I^{-1}(\hat{\boldsymbol{\theta}}_n)/n)$, a multivariate normal approximation, where $I^{-1}(\boldsymbol{\theta})$ is the inverse information matrix and $\hat{\boldsymbol{\theta}}_n$ is the MLE of $\boldsymbol{\theta}$. See, for example, Ferguson (1996, pp. 140-141). Then for large n, the posterior pdf is approximately the point mass at $\hat{\boldsymbol{\theta}}_n$, and $f_{Z|\boldsymbol{y}}(z|\boldsymbol{y}) \approx f(z|\hat{\boldsymbol{\theta}}_n)$. These heuristics suggest that the PIs based on the Bayesian predictive distribution and the PIs (3.3) and (3.4) will be similar for large n.

Section 3.2 describes some parametric models where the prediction intervals (3.3) and (3.4) are useful. Section 3.3 gives a simulation.

3.2 EXAMPLES

Suppose that $Y_1, ..., Y_n, Y_f$ are iid where Y_f is a future value and $\boldsymbol{Y}_n = (Y_1, ..., Y_n)^T$ is the data vector. If X is (left) truncated at d then $W = X|(X > d)$ has survival function $S_W(x) = \dfrac{S_X(x)}{S_X(d)}$ for $x > d$, and cumulative distribution function (cdf) $F_W(x) = 1 - S_W(x)$ for $x > d$. For insurance, losses are truncated if there is a deductible d.

Let Y_i = loss or time to event for ith person. $Y_i^* = T_i = \min(Y_i, Z_i)$ where Y_i and Z_i are independent and Z_i is the censoring random variable for the ith person (for time until event, Z_i is the time the ith person is lost to the study for any reason other than the time to event under study, often death). Let $\delta_i = I(Y_i \leq Z_i)$ so $\delta_i = 1$ if T_i is uncensored and $\delta_i = 0$ if T_i is censored. Alternatively, the censored data is $y_1, y_2+, y_3, ..., y_{n-1}, y_n+$ where y_i means the time was uncensored and y_i+ means the time was censored. For insurance, losses are right censored if there is a policy limit u: if the amount of the policy holder's loss exceeds u, then the benefit paid is u and the exact value of the loss for the policy holder is not recorded (the loss for the insurance company is the benefit paid).

Several random variables used in this section are briefly described next. We follow Klugman, Panjer, and Willmot (2008) closely.

$X \sim$ Gamma(α, θ) if X has probability density function (pdf)

$$f(x) = \frac{1}{\theta^\alpha \Gamma(\alpha)} x^{\alpha-1} e^{-x/\theta}$$

where α, θ, and x are positive. Then $E(X) = \alpha\theta$, and $V(X) = \alpha\theta^2$.

$X \sim$ Exponential(θ) if $X \sim$ Gamma$(\alpha = 1, \theta)$ or if $X \sim$ Weibull$(\theta, \tau = 1)$. Thus the pdf is $f(x) = \frac{1}{\theta} e^{-x/\theta}$ where $x, \theta > 0$. Then $E(X) = \theta$, and $V(X) = \theta^2$.

$X \sim$ Pareto(α, θ) if the pdf of X is $f(x) = \frac{\alpha\theta^\alpha}{(\theta + x)^{\alpha+1}}$ where α, θ, and x are positive. Then $E(X) = \frac{\theta}{\alpha - 1}$ for $\alpha > 1$, and $V(X) = \frac{\theta^2\alpha}{(\alpha-1)^2(\alpha-2)}$ for $\alpha > 2$.

$X \sim$ single parameter Pareto(α, θ) if the pdf of X is $f(x) = \frac{\alpha\theta^\alpha}{x^{\alpha+1}} I(x > \theta)$ where $\alpha > 0$ and θ is real. Note the **support** is $x > \theta$. Then $E(X) = \frac{\alpha\theta}{\alpha - 1}$ for $\alpha > 1$ and $V(X) = \frac{\alpha\theta^2}{\alpha - 2} - \left(\frac{\alpha\theta}{\alpha - 1}\right)^2$ for $\alpha > 2$.

$X \sim$ Weibull(θ, τ) if the pdf of X is $f(x) = \frac{\tau (x/\theta)^\tau e^{-(x/\theta)^\tau}}{x}$ where $\theta > 0$ and $\tau > 0$. Then $E(X^k) = \theta^k \Gamma(1 + k/\tau)$ for $k > -\tau$. $V(X) = E(X^2) - [E(X)]^2$.

Let X have a negative binomial NB(r,β) distribution where $\beta, r > 0$. Then the probability mass function (pmf) of X is $p_0 = (1 + \beta)^{-r}$, and for $k = 1, 2, ...,$

$$p_k = \frac{r(r+1)\cdots(r+k-1)\beta^k}{k!(1+\beta)^{r+k}} \text{ and } p_k = \frac{(k+r-1)!\beta^k}{k!(r-1)!(1+\beta)^{r+k}} \text{ for integer r.}$$

Then the expected value $E(X) = r\beta$ and the variance $V(X) = r\beta(1 + \beta)$. The parameterization

$$\left(\frac{1}{1+\beta}\right)^r \left(\frac{\beta}{1+\beta}\right)^k = \rho^r (1-\rho)^k$$

with $\rho = (1 + \beta)^{-1}$ is also used.

$X \sim$ Poisson(λ) if the pmf of X is $p_k = \frac{e^{-\lambda}\lambda^k}{k!}$ for $k = 0, 1, \ldots,$ where $\lambda > 0$. Then $E(X) = \lambda = V(X)$.

Example 1. Suppose that $Y_1, ..., Y_n, Y_f$ are iid Poisson(λ). Following Klugman, Panjer, and Willmot (2008, p. 412), if $W = Y_f$ and the prior distribution $\lambda \sim$ Gamma(α, θ), then the predictive distribution

$$W|\boldsymbol{Y}_n \sim NB\left(r = \alpha + \sum_{i=1}^n Y_i, \beta = \frac{\theta}{1 + n\theta}\right).$$

We could use PI (3.3) or (3.4) or generate $Y_1^*, ..., Y_B^*$ from the above predictive distribution and and apply the shorth interval (3.3). See Chen and Shao (1999). If $r_n\beta_n \to \lambda$ and $\beta_n \to 0$, then the negative binomial distribution converges in distribution to the Poisson(λ) random variable. See Agresti (2002, p. 560). Since $E(W|\boldsymbol{Y}_n) \approx V(W|\boldsymbol{Y}_n) \approx \overline{Y}_n$, which is the MLE $\hat{\lambda}$ of λ, the Bayesian and frequentist methods give similar results for large n.

Some MLE Formulas.

Let m = number of uncensored observations, c = number of censored observations, $n = m+c$, let d_i be the truncation point for each observation (0 if untruncated). Let x_i be the observation if uncensored or the censoring point (u_i) if censored. The following MLE formulas work if left truncation and right censoring are present or not.

a) EXP(θ): $\hat{\theta} = \dfrac{\sum_{i=1}^n (x_i - d_i)}{m}$.

b) Weibull fixed τ: $\hat{\theta} = \left(\dfrac{\sum_{i=1}^n (x_i^\tau - d_i^\tau)}{m} \right)^{1/\tau}$.

c) Pareto fixed θ: $\hat{\alpha} = \dfrac{-m}{\sum_{i=1}^n \ln\left(\frac{\theta+d_i}{\theta+x_i}\right)}$.

d) single parameter Pareto fixed θ: $\hat{\alpha} = \dfrac{-m}{\sum_{i=1}^n \ln\left(\frac{\max(\theta,d_i)}{x_i}\right)}$.

3.3 SIMULATIONS

R uses $Y \sim$ Exponential(λ) with $E(Y) = \theta = 1/\lambda$. The simulation generated $Y_i \sim$ Exponential(λ) and $Z_i \sim$ Exponential($\lambda = 0.1$) with $E(Z_i) = 10$. The MLE $\hat{\lambda} = 1/\hat{\theta}$ was computed. Then $Y_i^* \sim$ Exponential($\hat{\lambda}$) were computed for $i = 1, ..., B$. Then the 95% PI (3.4) was computed with $d = 1$. 5000 runs were used and the average length of the 5000 PIs was computed. The coverage is the percentage of times that the PIs contained the future value Y_f. Coverage between 0.94 and 0.96 suggest that the actual coverage is close to the nominal coverage 0.95. The population shorth is the interval $[0, -\theta \ln(0.05)] \approx [0, 2.9957/\lambda]$. Need n and B very large to get the average length close to the length of the population shorth.

```
source("http://parker.ad.siu.edu/Olive/survpack.txt")
args(exppisim)
```

```
function(n=100,nruns=100,B=1000,lam=1,clam=0.1,alpha=0.05)

exppisim(n=100,nruns=5000,B=1000,lam=1)

$lam

[1] 1

$mle

[1] 1.104127

$fullpicov

[1] 0.956

$fullpimenlen

[1] 3.284781
```

3.4 CONCLUSION

Since PIs (3.3) and (3.4) are for a parametric model, it is crucial to check that the parametric model is appropriate. In a similar application, Chen and Shao (1999) and Olive (2014, p. 364) used the shorth estimator to estimate Bayesian credible regions.

The simulations were done in *R*. See R Core Team (2016). The collection of Olive (2022) *R* functions *survpack*, available from (http://parker.ad.siu.edu/Olive/survpack.txt), has some useful functions for the inference. The functions `exppisim` and `mshpi` were used to make Table 3.1 and Table 3.2.

Table 3.1. 95% PIs for Right Censored EXP(λ) Data

n	λ	B	$3/\lambda$	length	coverage
50	1	1000	3	3.4442	0.9620
50	0.5	1000	6	6.8712	0.9574
50	2	1000	1.5	1.7143	0.9608
50	0.1	1000	30	35.0727	0.9650
50	10	1000	0.3	0.3425	0.9650
100	1	1000	3	3.2848	0.9560
100	0.5	1000	6	6.5807	0.9592
100	2	1000	1.5	1.6415	0.9632
100	0.1	1000	30	33.0942	0.9588
100	10	1000	0.3	0.3281	0.9546
1000	1	1000	3	3.1547	0.9508
1000	0.5	1000	6	6.3161	0.9568
1000	2	1000	1.5	1.5768	0.9530
1000	0.1	1000	30	31.5904	0.9542
1000	10	1000	0.3	0.3156	0.9640

Table 3.2. 95% PIs for Right Censored EXP(λ) Data

n	λ	B	$3/\lambda$	length	coverage
50	1	5000	3	3.3130	0.9632
50	0.5	5000	6	6.6383	0.9570
50	2	5000	1.5	1.6570	0.9578
50	0.1	5000	30	33.7727	0.9550
50	10	5000	0.3	0.3314	0.9614
100	1	5000	3	3.1814	0.9592
100	0.5	5000	6	6.3623	0.9552
100	2	5000	1.5	1.5918	0.9536
100	0.1	5000	30	32.2109	0.9618
100	10	5000	0.3	0.3180	0.9562
1000	1	5000	3	3.0665	0.9544
1000	0.5	5000	6	6.1398	0.9532
1000	2	5000	1.5	1.5337	0.9570
1000	0.1	5000	30	30.6792	0.9490
1000	10	5000	0.3	0.3070	0.9514

CHAPTER 4

KERNEL METHOD PROOFS

This chapter will use the kernel method to prove some results that are used in the life contingencies actuarial exams.

4.1 THE KERNEL METHOD

Using the fact that a probability density function integrates to 1 is often useful for integration. Similarly, a probability mass function (pmf) sums to 1. Notation such as $E(Y|\boldsymbol{\theta}) \equiv E_{\boldsymbol{\theta}}(Y)$ or $f_Y(y|\boldsymbol{\theta})$ is used to indicate that the formula for the expected value or pdf are for a family of distributions indexed by $\boldsymbol{\theta} \in \Theta$.

Following Olive (2014, pp. 12-13), the *kernel method* is a widely used technique for finding $E[g(Y)]$.

Definition. Let $f_Y(y)$ be the pdf or pmf of a random variable Y and suppose that $f_Y(y|\boldsymbol{\theta}) = c(\boldsymbol{\theta})k(y|\boldsymbol{\theta})$. Then $k(y|\boldsymbol{\theta}) \geq 0$ is the **kernel** of f_Y and $c(\boldsymbol{\theta}) > 0$ is the constant term that makes f_Y sum or integrate to one. Thus $\int_{-\infty}^{\infty} k(y|\boldsymbol{\theta})dy = 1/c(\boldsymbol{\theta})$ or $\sum_{y \in \mathcal{Y}} k(y|\boldsymbol{\theta}) = 1/c(\boldsymbol{\theta})$.

Often $E[g(Y)]$ is found using "tricks" tailored for a specific distribution. The word "kernel" means "essential part." Notice that if $f_Y(y)$ is a pdf, then $E[g(Y)] = \int_{-\infty}^{\infty} g(y)f(y|\boldsymbol{\theta})dy = \int_{\mathcal{Y}} g(y)f(y|\boldsymbol{\theta})dy$. Suppose that after algebra, it is found that

$$E[g(Y)] = a\ c(\boldsymbol{\theta}) \int_{-\infty}^{\infty} k(y|\boldsymbol{\tau})dy$$

for some constant a where $\boldsymbol{\tau} \in \Theta$ and Θ is the parameter space. Then the kernel method says that

$$E[g(Y)] = a\ c(\boldsymbol{\theta}) \int_{-\infty}^{\infty} \frac{c(\boldsymbol{\tau})}{c(\boldsymbol{\tau})} k(y|\boldsymbol{\tau})dy = \frac{a\ c(\boldsymbol{\theta})}{c(\boldsymbol{\tau})} \underbrace{\int_{-\infty}^{\infty} c(\boldsymbol{\tau})k(y|\boldsymbol{\tau})dy}_{1} = \frac{a\ c(\boldsymbol{\theta})}{c(\boldsymbol{\tau})}.$$

Similarly, if $f_Y(y)$ is a pmf, then

$$E[g(Y)] = \sum_{y: f(y) > 0} g(y)f(y|\boldsymbol{\theta}) = \sum_{y \in \mathcal{Y}} g(y)f(y|\boldsymbol{\theta})$$

where $\mathcal{Y} = \{y : f_Y(y) > 0\}$ is the support of Y. Suppose that after algebra, it is found that

$$E[g(Y)] = a\ c(\boldsymbol{\theta}) \sum_{y \in \mathcal{Y}} k(y|\boldsymbol{\tau})$$

for some constant a where $\boldsymbol{\tau} \in \Theta$. Then the kernel method says that

$$E[g(Y)] = a\, c(\boldsymbol{\theta}) \sum_{y \in \mathcal{Y}} \frac{c(\boldsymbol{\tau})}{c(\boldsymbol{\tau})} k(y|\boldsymbol{\tau}) = \frac{a\, c(\boldsymbol{\theta})}{c(\boldsymbol{\tau})} \underbrace{\sum_{y \in \mathcal{Y}} c(\boldsymbol{\tau}) k(y|\boldsymbol{\tau})}_{1} = \frac{a\, c(\boldsymbol{\theta})}{c(\boldsymbol{\tau})}.$$

The kernel method is often useful for finding $E[g(Y)]$, especially if $g(y) = y$, $g(y) = y^2$ or $g(y) = e^{ty}$. The kernel method is often easier than memorizing a trick specific to a distribution because the kernel method uses the same trick for every distribution: $\sum_{y \in \mathcal{Y}} f(y) = 1$ and $\int_{y \in \mathcal{Y}} f(y) dy = 1$. Of course sometimes tricks are needed to get the kernel $f(y|\boldsymbol{\tau})$ from $g(y) f(y|\boldsymbol{\theta})$. For example, complete the square for the normal (Gaussian) kernel.

Example. To use the kernel method to find the moment generating function of a gamma (ν, λ) distribution, note that

$$m(t) = E(e^{tY}) = \int_0^\infty e^{ty} \frac{y^{\nu-1} e^{-y/\lambda}}{\lambda^\nu \Gamma(\nu)} dy = \frac{1}{\lambda^\nu \Gamma(\nu)} \int_0^\infty y^{\nu-1} \exp[-y(\frac{1}{\lambda} - t)] dy.$$

The integrand is the kernel of a gamma (ν, η) distribution with

$$\frac{1}{\eta} = \frac{1}{\lambda} - t = \frac{1 - \lambda t}{\lambda} \quad \text{so} \quad \eta = \frac{\lambda}{1 - \lambda \mathrm{t}}.$$

Now

$$\int_0^\infty y^{\nu-1} e^{-y/\lambda} dy = \frac{1}{c(\nu, \lambda)} = \lambda^\nu \Gamma(\nu).$$

Hence

$$m(t) = \frac{1}{\lambda^\nu \Gamma(\nu)} \int_0^\infty y^{\nu-1} \exp[-y/\eta] dy = c(\nu, \lambda) \frac{1}{c(\nu, \eta)} =$$

$$\frac{1}{\lambda^\nu \Gamma(\nu)} \eta^\nu \Gamma(\nu) = \left(\frac{\eta}{\lambda}\right)^\nu = \left(\frac{1}{1 - \lambda t}\right)^\nu$$

for $t < 1/\lambda$.

4.2 LIFE CONTINGENCIES FORMULAS

Many of the following formulas can be found in Bowers et al. (1997), Camilli, Duncan, and London (2014), Cunningham, Herzog, and London (2008), Dickson, Hardy, and Waters (2020), and Weishaus (2010).

For life contingencies, the exponential(β) random variable = the gamma($\nu = 1, \lambda = 1/\beta$) random variable with pdf

$$f(x) = \beta \exp(-\beta x)\, I(x \geq 0) \text{ where } \beta > 0.$$

Then the expected value $E(X) = 1/\beta$, the variance $V(X) = 1/\beta^2$, and the cumulative distribution function $F(x) = 1 - \exp(-\beta x),\ x \geq 0$. Here $I(x \geq 0) = 1$ if $x \geq 0$ and $I(x \geq 0) = 0$, otherwise. The force of mortality $\mu(x) = \beta$ for $x > 0$. Often β is replaced by μ.

Let T_0 correspond to the lifetime of an object at birth or when the object is made. Let T_x be the time until failure of the object given that the object has survived to time $x > 0$. If $T_0 \sim EXP(\mu)$, then $T_x \sim EXP(\mu)$. This result is known as the memoryless property of the exponential distribution. We use $\stackrel{E}{=}$ when the exponential RV is used.

Let $t > 0$. Let $f_x(t) = f_{T_x}(t)$, $F_x(t) = F_{T_x}(t)$, $S_x(t) = S_{T_x}(t)$ and $\mu_x(t) = \mu_{T_x}(t)$.

a)

$${}_tp_x = \frac{S_0(x+t)}{S_0(x)} = 1 - {}_tq_x = P(T_x > t) = P(T_0 > x + t | T_0 > x) = S_x(t)$$

$$= \exp(-\int_x^{x+t} \mu_r \, dr) = \exp(-\int_0^t \mu_{x+s} \, ds)$$

Note that $S_0(x + t) = S_0(x) S_x(t)$.

b)

$${}_tq_x = 1 - {}_tp_x = 1 - \frac{S_0(x+t)}{S_0(x)} = P(T_x \leq t) = P(T_0 \leq x + t | T_0 > x) = F_x(t)$$

c)

$$f_x(t) = \frac{f_0(x+t)}{S_0(x)} = {}_tp_x \, \mu_{x+t} = \frac{d}{dt} F_x(t) = -\frac{d}{dt} S_x(t)$$

d)

$$\mu_{x+t} = \frac{f_0(x+t)}{S_0(x+t)} = \mu_0(x+t) = \mu_x(t)$$

Suppose (x) buys insurance and dies at $t > 0$ years from purchase so $T = T_x = t$. Suppose a unit payment (eg of \$100000, \$500000 or \$1000000) is made. Then $v = \dfrac{1}{1+i} = e^{-\delta}$ and $\delta = \log(1 + i) = -\log(v)$. Often use $v^t = e^{-\delta t}$ and $v^{2t} = e^{-2\delta t}$.

The rule of moments for $b_t \in \{0, 1\}$ (unit payment insurance) is if $E[\,\overline{Z}\,] = \overline{A} = g(\delta)$, then $E[(\overline{Z})^j] = {}^j\overline{A} = g(j\delta)$. This rule is usually used for $j = 2$. Often Z is used instead of $\overline{Z}$.

I) Continuous *whole life insurance* makes unit payment at time $t = k$ with $v_t = v^t, t \geq 0$ and $b_t = 1, t \geq 0$. Then $z_t = b_t v_t = v^t, t \geq 0$. The present value random variable $\overline{Z}_x = z_T = v^T$. Then the actuarial present value APV = EPV = NSP =

$$\overline{A}_x = E(\overline{Z}_x) = E(v^T) = E(e^{-\delta T}) = \int_0^\infty v^t f_T(t)\,dt = \int_0^\infty e^{-\delta t} f_T(t)\,dt = \int_0^\infty v^t\ {}_tp_x\,\mu_{x+t}\,dt, \text{ and}$$

$${}^2\overline{A}_x = E[(\overline{Z}_x)^2] = E[(v^T)^2] = E(e^{-2\delta T}) = \int_0^\infty v^{2t} f_T(t)\,dt = \int_0^\infty e^{-2\delta t} f_T(t)\,dt = \int_0^\infty v^{2t}\ {}_tp_x\,\mu_{x+t}\,dt.$$

The moment generating function of a nonnegative random variable T is $m_T(z) = E(e^{zT}) = \int_0^\infty e^{zt} f_T(t)\,dt$ provided that $m_T(z)$ exists in a neighborhood of $z = 0$. Thus $\overline{A}_x = m_T(-\delta)$ and ${}^2\overline{A}_x = m_T(-2\delta)$. A problem with this formula is $T = T_x$ usually does not have a nice distribution or moment generating function even if T_0 has a nice distribution.

II) Continuous *n year term insurance* makes unit payment at time $t > 0$ only if $t \leq n$, otherwise no payment is made. Now $v_t = v^t, t \geq 0$,

$$b_t = \begin{cases} 1, & t \leq n \\ 0, & t > n, \end{cases} \qquad z_t = b_t v_t = \begin{cases} v^t, & t \leq n \\ 0, & t > n, \end{cases} \quad \text{and} \quad \overline{Z}^{\,1}_{x:\overline{n}|} = \begin{cases} v^{T_x}, & T \leq n \\ 0, & T > n. \end{cases}$$

Then the actuarial present value APV = EPV = NSP =

$$\overline{A}^{\,1}_{x:\overline{n}|} = E(\overline{Z}^{\,1}_{x:\overline{n}|}) = \int_0^n e^{-\delta t} f_T(t)\,dt = \int_0^n v^t f_T(t)\,dt = \int_0^n v^t\ {}_tp_x\,\mu_{x+t}\,dt, \text{ and}$$

$${}^2\overline{A}^{\,1}_{x:\overline{n}|} = E[(\overline{Z}^{\,1}_{x:\overline{n}|})^2] = \int_0^n e^{-2\delta t} f_T(t)\,dt = \int_0^n v^{2t} f_T(t)\,dt = \int_0^n v^{2t}\ {}_tp_x\,\mu_{x+t}\,dt.$$

The 1 above the x means unit benefit is payable after (x) dies if death is not after time n.

III) Continuous *n year deferred insurance* makes unit payment at time $t > 0$ only if $t > n$, otherwise no payment is made. Now $v_t = v^t, t \geq 0$,

$$b_t = \begin{cases} 0, & t \leq n \\ 1, & t > n \end{cases} \quad \text{and} \quad z_t = b_t v_t = \begin{cases} 0, & t \leq n \\ v^t, & t > n. \end{cases}$$

The present value random variable

$$ {}_{n|}\overline{Z}_x = \begin{cases} 0, & T \le n \\ v^T, & T > n. \end{cases} $$

Then the actuarial present value APV = EPV = NSP =

$$ {}_{n|}\overline{A}_x = E({}_{n|}\overline{Z}_x) = \int_n^\infty e^{-\delta t} f_T(t)\, dt = \int_n^\infty v^t f_T(t)\, dt = \int_n^\infty v^t\ {}_tp_x\, \mu_{x+t}\, dt, \text{ and} $$

$$ {}^2\ {}_{n|}\overline{A}_x = E[({}_{n|}\overline{Z}_x)^2] = \int_n^\infty e^{-2\delta t} f_T(t)\, dt = \int_n^\infty v^{2t} f_T(t)\, dt = \int_n^\infty v^{2t}\ {}_tp_x\, \mu_{x+t}\, dt. $$

IV) Discrete = continuous *n year pure endowment insurance* makes unit payment at time n only if $t > n$, otherwise no payment is made. Now

$$ v_t = \begin{cases} v^t, & t \le n \\ v^n, & t > n, \end{cases} \qquad b_t = \begin{cases} 0, & t \le n \\ 1, & t > n \end{cases} \qquad \text{and } \mathrm{z_t = b_t v_t} = \begin{cases} 0, & t \le n \\ v^n, & t > n. \end{cases} $$

The present value random variable

$$ Z_{\overset{\;1}{x:\overline{n}|}} = \begin{cases} 0, & T_x \le n \\ v^n, & T_x > n. \end{cases} $$

Then the actuarial present value APV = EPV = NSP =

$$ A_{\overset{\;1}{x:\overline{n}|}} = E(Z_{\overset{\;1}{x:\overline{n}|}}) = {}_nE_x = v^n P(T_x > n) = v^n \int_n^\infty f_x(t)\, dt = v^n \int_n^\infty {}_tp_x\, \mu_{x+t}\, dt = v^n\ {}_np_x $$

$(= v^n P(K_x \ge n) = v^n \sum_{k=n}^\infty P(K_x = k) = v^n S_x(n) = e^{-\delta n} S_x(n)$ and

$$ {}^2A_{\overset{\;1}{x:\overline{n}|}} = E[(Z_{\overset{\;1}{x:\overline{n}|}})^2] = v^{2n} P(T_x > n) = v^{2n} \int_n^\infty f_x(t)\, dt = v^{2n} \int_n^\infty {}_tp_x\, \mu_{x+t}\, dt = v^{2n}\ {}_np_x $$

$= v^{2n} P(K_x \ge n) = v^{2n} \sum_{k=n}^\infty P(K_x = k) = v^{2n} S_x(n) = e^{-2\delta n} S_x(n)$. The 1 above the $\overline{n}|$ means unit benefit is payable after (x) dies if death is after time n.

Also $V(Z_{\overset{\;1}{x:\overline{n}|}}) = v^{2n}\ {}_np_x\ {}_nq_x$.

V) Continuous *n year endowment life insurance* is a term insurance plus a pure endowment insurance, and makes unit payment at time $t > 0$ if $t < n$ and at time n if $t > n$. Then $b_t = 1, t \geq 0$ and

$$v_t = \begin{cases} v^t, & t \leq n \\ v^n, & t > n \end{cases} \quad \text{and } z_t = b_t v_t = \begin{cases} v^t, & t \leq n \\ v^n, & t > n. \end{cases}$$

The present value random variable

$$\overline{Z}_{x:\overline{n}|} = \begin{cases} v^T, & T \leq n \\ v^n, & T > n. \end{cases}$$

Note that the n year endowment present value random variable $\overline{Z}_{x:\overline{n}|} = \overline{Z}^{1}_{x:\overline{n}|} + Z_{x:\overline{n}|}^{\ \ 1}$, the sum of the n year term and n year pure endowment present value RVs.

Then the actuarial present value APV = EPV = NSP =

$$\overline{A}_{x:\overline{n}|} = E[\overline{Z}_{x:\overline{n}|}] = \overline{A}^{1}_{x:\overline{n}|} + A_{x:\overline{n}|}^{\ \ 1} = \int_0^n v^t f_T(t)\, dt + v^n P(T > n) = \int_0^n v^t\, {}_tp_x\, \mu_{x+t}\, dt + v^n\, {}_np_x.$$

Similarly, $[\overline{Z}_{x:\overline{n}|}]^2 = [\overline{Z}^{1}_{x:\overline{n}|}]^2 + [Z_{x:\overline{n}|}^{\ \ 1}]^2$ and ${}^2\overline{A}_{x:\overline{n}|} = {}^2\overline{A}^{1}_{x:\overline{n}|} + {}^2A_{x:\overline{n}|}^{\ \ 1}$

$$= \int_0^n v^{2t} f_T(t)\, dt + v^{2n} P(T_x > n) = \int_0^n v^{2t}\, {}_tp_x\, \mu_{x+t}\, dt + v^{2n}\, {}_np_x.$$

VI) A continuous n year deferred m year term insurance pays 1 unit at time t only if $n < t \leq n + m$ with $b_t = 0$ for $t \leq n$ and $t > n + m$ and $b_t = 1$ for $n < t < n + m$. Then $z_t = b_t v_t$ and $\overline{Z}_T = v_T = e^{-\delta t}$ for $n < T \leq n + m$ and $Z_T = 0$ for $T < n$ or $T > n + m$. Then

$$E(\overline{Z}_T) = {}_{n|m}\overline{A}_x = {}_{n|}\overline{A}^{1}_{x:\overline{m}|} = \overline{A}_x({}_nE_x - {}_{n+m}E_x) = \int_n^{n+m} e^{-\delta t} f_T(t)dt$$

$$\overset{E}{=} \frac{\mu}{\mu + \delta}[e^{-n(\mu+\delta)} - e^{-(n+m)(\mu+\delta)}].$$

VII) Continuous increasing whole life insurance pays t units at time t and has $v_t = v^t = e^{-\delta t}$ and $b_t = t$ for $t \geq 0$. So $z_t = b_t v_t = t v^t$ and the present value RV $\overline{Z} = \overline{B}_x = z_{T_x} = T_x v^{T_x}$. Hence the APV $= E(\overline{Z}) = E[\overline{B}_x] = (\overline{I}\,\overline{A})_x = \int_0^\infty t e^{-\delta t} f_x(t) dt = \int_0^\infty t v^t\, {}_tp_x\, \mu_{x+t}\, dt.$

VIII) A continuous whole life annuity makes a continuous payment at an annual rate of 1 unit per year as long as (x) survives. The present value RV

$$\overline{Y}_x = \overline{a}_{\overline{T_x}|} = \frac{1 - v^{T_x}}{\delta} = \frac{1 - \overline{Z}_x}{\delta}.$$

The APV is $\overline{a}_x = E(\overline{Y}_x) = \int_0^\infty v^t \, {}_tp_x dt = \int_0^\infty e^{-\delta t} S_x(t)dt$ where $S_x(t) = S_0(x+t)/S_0(x) = S_{T_x}(t)$. Note the δ in the denominator of the continuous annuity. $V(\overline{Y}_x) = \frac{V(\overline{Z}_x)}{\delta^2} = \frac{{}^2\overline{A}_x - (\overline{A}_x)^2}{\delta^2}$.

$$\overline{Y}_x = \frac{1 - \overline{Z}_x}{\delta}.$$

$E[\overline{Y}_x] = \overline{a}_x = \int_0^\infty e^{-\delta t} S_T(t)dt \stackrel{E}{=} \frac{1}{\mu + \delta}$. $V(\overline{Y}_x) = \frac{V(\overline{Z}_x)}{\delta^2} = \frac{{}^2\overline{A}_x - (\overline{A}_x)^2}{\delta^2}$.

Multiple life functions consider failure or survival of a *status* of multiple lives. Insurance is payable when the status fails. Annuities are payable as long as the status survives. For 2 life functions the x and y are separated by a colon. So think of (xy) as $(x{:}y)$, and $(\overline{xy})$ as $(\overline{x:y})$. Notation $x+n : y+n$ is also used. Let $T_x \perp\!\!\!\perp T_y$ mean that T_x and T_y are independent. Usually assume $T_{x_1} \perp\!\!\!\perp T_{x_2} \perp\!\!\!\perp ... \perp\!\!\!\perp T_{x_k}$.

A **joint life status** for (xy) fails as soon as x or y dies. Let $T_{xy} = \min(T_x, T_y)$ = time until 1st death. A two life **last survivor status** for $(\overline{xy})$ fails after both x and y die. Let $T_{\overline{xy}} = \max(T_x, T_y)$ = time until 2nd death. Then $T_{xy} + T_{\overline{xy}} = T_x + T_y$.

If $T_x \sim EXP(\mu_x) \perp\!\!\!\perp T_y \sim EXP(\mu_y)$, then $T_{xy} = \min(T_x, T_y) \sim EXP(\mu_x + \mu_y)$.

Let $T_{x_1}, ..., T_{x_m}$ be independent EXP(μ_i) RVs. Let $u = (x_1 \cdots x_m)$ or $u = x_1 \cdots x_m$. Then $T = T_u = T_{x_1 \cdots x_m} = \min(T_{x_1}, ..., T_{x_m}) \sim EXP(\sum_{i=1}^m \mu_i)$. Then $\mu_T(t) = \sum_{i=1}^m \mu_i$, $S_T(t) = \exp(-t \sum_{i=1}^m \mu_i)$, $\overset{o}{e}_u = E(T) = 1/(\sum_{i=1}^m \mu_i)$ and $V(T) = 1/(\sum_{i=1}^m \mu_i)^2$.

IX) For whole life insurance, $\overline{A}_u = E[\overline{Z}_u] = \frac{\sum_{i=1}^m \mu_i}{\delta + \sum_{i=1}^m \mu_i}$, and ${}^2\overline{A}_u = E[(\overline{Z}_u)^2] = \frac{\sum_{i=1}^m \mu_i}{2\delta + \sum_{i=1}^m \mu_i}$.

X) For a whole life annuity, $\overline{a}_u = E[\overline{Y}_u] = \frac{1}{\delta + \sum_{i=1}^m \mu_i}$, and $V[\overline{Y}_u] = \frac{{}^2\overline{A}_u - (\overline{A}_u)^2}{\delta^2}$.

4.3 SOME PROOFS

Let $T \sim EXP(\mu)$ with $S(t) = e^{-\mu t}$ for $t > 0$. Then $E(T) = \int_0^\infty t\mu e^{-\mu t} dt = \int_0^\infty e^{-\mu t} dt = 1/\mu$. So $\int_0^\infty tDe^{-t(D)} dt = \int_0^\infty e^{-t(D)} dt = 1/D$ for $D > 0$. Use $\overset{E}{=}$ when the exponential RV is used.

0) Continuous whole life insurance with the exponential(μ) distribution often has $\overline{Z} = b_t \, v_t = e^{\theta t} e^{-\delta t}$ where $b_t = e^{\theta t}$ and $v_t = e^{-\delta t}$. Then $E[\overline{Z}] \overset{E}{=} \dfrac{\mu}{\mu + \delta - \theta}$ provided $\mu + \delta - \theta > 0$. Also $E[(\overline{Z})^j] \overset{E}{=} \dfrac{\mu}{\mu + \delta j - \theta j}$ provided $\mu + \delta j - \theta j > 0$.

Proof: Now $\displaystyle\int_0^\infty \mu e^{-\mu t} dt = 1$ so $\displaystyle\int_0^\infty e^{-\mu t} dt = 1/\mu$ if $\mu > 0$. Hence $E[\overline{Z}] = \displaystyle\int_0^\infty b_t e^{-\delta t} \mu e^{-\mu t} dt = \int_0^\infty e^{\theta t} e^{-\delta t} \mu e^{-\mu t} dt = \mu \int_0^\infty e^{-t(\mu + \delta - \theta)} dt = \frac{\mu}{\mu + \delta - \theta}$ provided $\mu + \delta - \theta > 0$. Also $E[(\overline{Z})^j] = \displaystyle\int_0^\infty [b_t e^{-\delta t}]^j \mu e^{-\mu t} dt = \int_0^\infty e^{\theta j t} e^{-\delta j t} \mu e^{-\mu t} dt = \mu \int_0^\infty e^{-t(\mu + \delta j - \theta j)} dt = \frac{\mu}{\mu + \delta j - \theta j}$ provided $\mu + \delta j - \theta j > 0$. □

Notes: i) If $b_t = c e^{\theta t}$ and $\overline{Z} = b_T v_T$, then $E[\overline{Z}^j] = E[(b_T v_T)^j] = c^j E[(e^{\theta T} v_T)^j]$. So multiply $c = 1$ formulas by c^j. Usually want $j = 1, 2$.

ii) If $T = T_x$ has moment generating function $m_T(z)$, then $E(\overline{Z}^j) = m_T(\theta j - \delta j) = m_T([\theta - \delta] j)$.

I) Continuous whole life insurance: special case of 0) with $\theta = 0$.

$\overline{Z}_x = e^{-\delta T}$. $\overline{A}_x = E(\overline{Z}_x) = E(e^{-\delta T}) \overset{E}{=} \dfrac{\mu}{\mu + \delta}$, and ${}^2\overline{A}_x = E[(\overline{Z}_x)^2] = E(e^{-2\delta T}) \overset{E}{=} \dfrac{\mu}{\mu + 2\delta}$. $V(\overline{Z}_x) = {}^2\overline{A}_x - (\overline{A}_x)^2$.

Proof: Note that $\theta = 0$ corresponds to unit payment, $b_t = e^{0t} = 1$. $E[\overline{Z}_x] = \displaystyle\int_0^\infty e^{-\delta t} \mu e^{-\mu t} dt = \mu \int_0^\infty e^{-t(\mu + \delta)} dt = \frac{\mu}{\mu + \delta}$ provided $\mu + \delta > 0$. Also $E[(\overline{Z}_x)^j] = \displaystyle\int_0^\infty e^{-\delta j t} \mu e^{-\mu t} dt = \mu \int_0^\infty e^{-t(\mu + \delta j)} dt = \frac{\mu}{\mu + \delta j}$ provided $\mu + \delta j > 0$. □

Often $\int_0^\infty$ is replaced by $\int_a^b$. Thus $E[\overline{Z}] = \displaystyle\int_a^b b_t e^{-\delta t} \mu e^{-\mu t} dt = \int_a^b e^{\theta t} e^{-\delta t} \mu e^{-\mu t} dt = \mu \int_a^b e^{-t(\mu + \delta - \theta)} dt = \frac{\mu}{\mu + \delta - \theta} [e^{-a(\mu + \delta - \theta)} - e^{-b(\mu + \delta - \theta)}]$ provided $\mu + \delta - \theta > 0$. And $E[(\overline{Z})^j] = \displaystyle\int_a^b [b_t e^{-\delta t}]^j \mu e^{-\mu t} dt = \int_a^b e^{\theta j t} e^{-\delta j t} \mu e^{-\mu t} dt = \mu \int_a^b e^{-t(\mu + \delta j - \theta j)} dt =$

$$\frac{\mu}{\mu + \delta j - \theta j} [e^{-a(\mu + \delta j - \theta j)} - e^{-b(\mu + \delta j - \theta j)}]$$

provided $\mu + \delta j - \theta j > 0$.

If $D > 0$, $\int_0^n De^{-tD}dt = 1 - e^{-nD}$, $\int_n^\infty De^{-tD}dt = e^{-nD}$, $\int_0^n e^{-tD}dt = \frac{1}{D}\int_0^n De^{-tD}dt = \frac{1}{D}[1 - e^{-nD}]$, and $\int_n^\infty e^{-tD}dt = \frac{1}{D}\int_n^\infty De^{-tD}dt = \frac{1}{D}\, e^{-nD}$.

II) Continuous *n year term insurance*: $\overline{A}^1_{x:\overline{n}|} = E(\overline{Z}^1_{x:\overline{n}|}) \overset{E}{=} \int_0^n e^{-\delta t}\mu e^{-\mu t}\, dt = \mu\int_0^n e^{-t(\mu+\delta)}dt = \frac{\mu}{\mu+\delta}[1 - e^{-n(\mu+\delta)}]$ and ${}^2\overline{A}^1_{x:\overline{n}|} = E[(\overline{Z}^1_{x:\overline{n}|})^2] \overset{E}{=} \int_0^n e^{-2\delta t}\mu e^{-\mu t}\, dt = \mu\int_0^n e^{-t(\mu+2\delta)}dt = \frac{\mu}{\mu+2\delta}[1 - e^{-n(\mu+2\delta)}]$.

III) Continuous *n year deferred insurance*:

$$_{n|}\overline{A}_x = E(_{n|}\overline{Z}_x) \overset{E}{=} \int_n^\infty e^{-\delta t}\mu e^{-\mu t}\, dt = \mu\int_n^\infty e^{-t(\mu+\delta)}dt = \frac{\mu}{\mu+\delta}[e^{-n(\mu+\delta)}]$$

and

$$^2\,{}_{n|}\overline{A}_x = E[(_{n|}\overline{Z}_x)^2] \overset{E}{=} \int_n^\infty e^{-2\delta t}\mu e^{-\mu t}\, dt = \mu\int_n^\infty e^{-t(\mu+2\delta)}dt = \frac{\mu}{\mu+2\delta}[e^{-n(\mu+2\delta)}].$$

IV) Discrete = continuous *n year pure endowment insurance*:

$$A_{\underset{x:\overline{n}|}{\ \ 1}} = E(Z_{\underset{x:\overline{n}|}{\ \ 1}}) = {}_nE_x = v^n P(T_x > n) = e^{-\delta n} S_x(n) \overset{E}{=} e^{-\delta n}e^{-\mu n} = e^{-n(\mu+\delta)}$$

and

$$^2A_{\underset{x:\overline{n}|}{\ \ 1}} = E[(Z_{\underset{x:\overline{n}|}{\ \ 1}})^2] = v^{2n} P(T_x > n) = e^{-2\delta n} S_x(n) \overset{E}{=} e^{-2\delta n}e^{-\mu n} = e^{-n(\mu+2\delta)}.$$

Note: If $S_x(t)$ is easy to derive, then the above quantities can be obtained. Hence the exponential distribution does not need to be used in hand calculations.

V) A continuous n year endowment life insurance with the exponential(μ) distribution has $\overline{A}_{x:\overline{n}|} = \overline{A}^1_{x:\overline{n}|} + A_{\underset{x:\overline{n}|}{\ \ 1}} = \overline{A}^1_{x:\overline{n}|} + {}_nE_x = \overline{A}_x + {}_nE_x(1 - \overline{A}_{x+n}) = \overline{A}_x + A_{\underset{x:\overline{n}|}{\ \ 1}}(1 - \overline{A}_{x+n}) \overset{E}{=} \frac{\mu}{\mu+\delta} + (e^{-\delta n}e^{-\mu n})(1 - \frac{\mu}{\mu+\delta}) = \frac{\mu}{\mu+\delta} + e^{-n(\mu+\delta)}\frac{\delta}{\mu+\delta} = \frac{\mu + \delta e^{-n(\mu+\delta)}}{\mu+\delta}$.

VI) A continuous n year deferred m year term insurance with the exponential(μ) distribution has $E(\overline{Z}_T) = {}_{n|m}\overline{A}_x = {}_{n|}\overline{A}^1_{x:\overline{m}|} = \overline{A}_x({}_nE_x - {}_{n+m}E_x) \overset{E}{=} \int_n^{n+m} e^{-\delta t}\mu e^{-\mu t}dt = \mu\int_n^{n+m} e^{-t(\mu+\delta)}dt = \frac{\mu}{\mu+\delta}[e^{-n(\mu+\delta)} - e^{-(n+m)(\mu+\delta)}]$.

VII) Continuous increasing whole life insurance: If $T \sim EXP(\mu)$, then $(\overline{I}\,\overline{A})_x \overset{E}{=} \frac{\mu}{(\mu+\delta)^2}$.

VIII) A continuous whole life annuity makes a continuous payment at an annual rate of 1 unit per year with the exponential(μ) distribution has $\overline{a}_x = E(\overline{Y}_x) \overset{E}{=} \frac{1}{\mu+\delta}$.

Proof: $\overline{a}_x = E(\overline{Y}_x) = \int_0^\infty v^t\ {}_tp_x dt \overset{E}{=} \int_0^\infty e^{-\delta t} e^{-\mu t} dt = \int_0^\infty e^{-t(\mu+\delta)} dt = \frac{1}{\mu+\delta}$.

□

Proofs for IX) and X):

For a multiple whole life insurance with $T_{x_1}, ..., T_{x_m}$ as independent EXP(μ_i) RVs and $u = (x_1 \cdots x_m)$ or $u = x_1 \cdots x_m$ with $T = T_u = T_{x_1 \cdots x_m} = \min(T_{x_1}, ..., T_{x_m}) \sim EXP(\sum_{i=1}^m \mu_i)$, we have $\overline{A}_u = E[\overline{Z}_u] = \int_0^\infty e^{-\delta t}(\sum_{i=1}^m \mu_i) e^{-(\sum_{i=1}^m \mu_i)t} dt = \sum_{i=1}^m \mu_i \int_0^\infty e^{-t(\delta + \sum_{i=1}^m \mu_i)} dt = \frac{\sum_{i=1}^m \mu_i}{\delta + \sum_{i=1}^m \mu_i}$, and ${}^2\overline{A}_u = E[(\overline{Z}_u)^2] = \int_0^\infty e^{-(2\delta)t}(\sum_{i=1}^m \mu_i) e^{-(\sum_{i=1}^m \mu_i)t} dt = \sum_{i=1}^m \mu_i \int_0^\infty e^{-t(2\delta + \sum_{i=1}^m \mu_i)} dt = \frac{\sum_{i=1}^m \mu_i}{2\delta + \sum_{i=1}^m \mu_i}$. For a multiple whole life annuity, $\overline{a}_u = E[\overline{Y}_u] = \int_0^\infty e^{-\delta t} e^{-(\sum_{i=1}^m \mu_i)t} dt = \int_0^\infty e^{-t(\delta + \sum_{i=1}^m \mu_i)} dt = \frac{1}{\delta + \sum_{i=1}^m \mu_i}$ by using the kernel method. □

REFERENCES

[1] Agresti, A. (2002), *Categorical Data Analysis*, 2nd ed., Wiley, Hoboken, NJ.

[2] Bowers, N.L., Gerber, H.U., Hickman, J.C., Jones, D.A. and Nesbitt, C.J. (1997), *Actuarial Mathematics*, 2nd ed., ACTEX Publications, Winsted, CT.

[3] Camilli, S.J., Duncan, I., and London, R.L. (2014), *Models for Quantifying Risk*, 6th ed. ACTEX Publications, Winsted, CT.

[4] Chen, M.H., and Shao, Q.M. (1999), "Monte Carlo Estimation of Bayesian Credible and HPD Intervals, *Journal of Computational and Graphical Statistics,* 8, 69-92.

[5] Cunningham, R.J., Herzog, T.N., and London, R.L. (2008), *Models for Quantifying Risk*, 3rd ed., ACTEX Publications, Winsted, CT.

[6] Dickson, D.C.M., Hardy, M.R., and Waters, H.R. (2020), *Actuarial Mathematics for Life Contingent Risks*, 3rd ed., Cambridge University Press, Cambridge, UK.

[7] Ferguson, T.S. (1996), *A Course in Large Sample Theory*, Chapman & Hall, New York, NY.

[8] Frey, J. (2013), "Data-Driven Nonparametric Prediction Intervals," *Journal of Statistical Planning and Inference*, 143, 1039-1048.

[9] Geisser, S. (1993), *Predictive Inference: an Introduction*, Chapman and Hall, New York, NY.

[10] Gunst, R.F., and Mason, R.L. (1980), *Regression Analysis and Its Application,* Marcel Dekker, New York, NY.

[11] Hall, P., Peng, L., and Tajvidi, N. (1999), "On Prediction Intervals Based on Predictive Likelihood or Bootstrap Methods," *Biometrika*, 86, 871-880.

[12] Hall, P., and Rieck, A. (2001), "Improving Coverage Accuracy of Nonparametric Prediction Intervals," *Journal of the Royal Statistical Society, B,* 63, 717-725.

[13] Hoerl, A.E., and Kennard, R. (1970), "Ridge Regression: Biased Estimation for Nonorthogonal Problems," *Technometrics*, 12, 55-67.

[14] Kellison, S.G. and London, R.L. (2011), *Risk Models and Their Estimation*, ACTEX Publications, Winsted, CT.

[15] Kibria, B.M.G., and Lukman, A.F. (2020), "A New Ridge-Type Estimator for the Linear

Regression Model: Simulations and Applications," *Scientifica*, 2020, online.

[16] Klugman, S.A.; Panjer, H.H.; and Willmot, G.E., (2008), *Loss Models: From Data to Decisions*, 3rd ed., Wiley, New York, NY.

[17] Knight, K., and Fu, W.J. (2000), "Asymptotics for Lasso-Type Estimators," *The Annals of Statistics*, 28, 1356–1378.

[18] Kurnaz, F.S., and Akay, K.U. (2015), "A New Liu-Type Estimator," *Statistical Papers*, 56, 495-517.

[19] Lawless, J.F., and Fredette, M. (2005), "Frequentist Prediction Intervals and Predictive Distributions," *Biometrika* 92, 529-542.

[20] Li, Y., and Yang, H. (2012), "A New Liu-Type Estimator in Linear Regression Model," *Statistical Papers*, 53, 427-437.

[21] Liu, K. (1993), "A New Class of Biased Estimate in Linear Regression," *Communications in Statistics: Theory and Methods* 22, 393-402

[22] Liu, K. (2003), "Using Liu-Type Estimator to Combat Collinearity," *Communications in Statistics: Theory and Methods*, 32, 1009-1020.

[23] Olive, D.J. (2007), "Prediction Intervals for Regression Models," *Computational Statistics & Data Analysis,* 51, 3115-3122.

[24] Olive, D.J. (2013), "Asymptotically Optimal Regression Prediction Intervals and Prediction Regions for Multivariate Data," *International Journal of Statistics and Probability*, 2, 90-100.

[25] Olive, D.J. (2014), *Statistical Theory and Inference*, Springer, New York, NY.

[26] Olive, D.J. (2022), *Survival Analysis*, online course notes, see (http://parker.ad.siu.edu/Olive/survbk.htm).

[27] Olive, D.J., Rathnayake, R.C., and Haile, M.G. (2021), "Prediction Intervals for GLMs, GAMs, and Some Survival Regression Models," *Communications in Statistics: Theory and Methods*, to appear.

[28] Özkale, M.R., and Kaçiranlar, S. (2007), "The Restricted and Unrestricted Two-Parameter Estimators," *Communications in Statistics: Theory and Methods*, 36, 2707-2725.

[29] Pelawa Watagoda, L.C.R., and Olive, D.J. (2021), "Comparing Six Shrinkage Estimators With Large Sample Theory and Asymptotically Optimal Prediction Intervals," *Statistical Papers,* 62, 2407-2431. See (http://parker.ad.siu.edu/Olive/pppicomp.pdf).

[30] R Core Team (2016), "R: a Language and Environment for Statistical Computing," R Foundation for Statistical Computing, Vienna, Austria, (www.R-project.org).

[31] Sakallioğlu, S., and Kaçiranlar, S. (2008), "A New Biased Estimator Based on Ridge Estimation," *Statistical Papers*, 49, 669-689.

[32] Sen, P.K., and Singer, J.M. (1993), *Large Sample Methods in Statistics: an Introduction with Applications,* Chapman & Hall, New York, NY.

[33] Tibshirani, R. (1996), "Regression Shrinkage and Selection Via the Lasso," *Journal of the Royal Statistical Society, B*, 58, 267-288.

[34] Ueki, M. and Fueda, K. (2007), "Adjusting Estimative Prediction Limits," *Biometrika,* 94, 509-511.

[35] Weishaus, A. (2010), *ASM Study Manual for SOA Exam MLC*, 10th ed., Actuarial Study Materials, Westbury, NY. (See www.studymanuals.com.)

[36] Yang, H., and Chang, X. (2010), "A New Two-Parameter Estimator in Linear Regression," *Communications in Statistics: Theory and Methods*, 39, 923-934.

VITA

Graduate School
Southern Illinois University Carbondale

Yu Jin

jinyu85@siu.edu
jinyu85@yahoo.com

Southern Illinois University Carbondale
Master of Science, Mathematics, May 2010

Dissertation Paper Title:
Some Topics Relevant to Actuarial Mathematics

Major Professor: Dr. David Olive