Module 1

Probability Theory

Probability theory is the backbone of actuarial science and plays a key role in most topics covered in the ACIA Capstone Exam. Although you will not face direct questions on probability theory itself, its applications are everywhere – from understanding financial risks to modeling life events like mortality.

Probability helps actuaries make informed decisions by breaking complex problems into manageable components. For example, life insurance calculations are heavily based on survival models, which are grounded in probability. Actuaries use these models to predict how long policyholders will live and to price products accordingly. These predictions are essential for developing reliable financial solutions that secure people's futures.

Probability theory is also practical when working on predictive analytics, a significant part of the exam. Creating models that forecast trends, such as insurance claims or investment returns, requires a strong foundation in probability. Understanding random variables and distributions helps you assess risks and make accurate predictions.

In this module, we will review the basics of probability theory and build a solid foundation for the other topics. Although not directly tested, mastering these concepts is essential to understand the applications that appear throughout the exam.

This page is intentionally left blank.

Contents

1.1 Probability

*****FROM THE CIA EDUCATION SYLLABUS*****

A1.1 Probability

a. Define set functions, sample space, and events. Define probability as a set function on a collection of events and state the basic axioms of probability.

b. Calculate probabilities of mutually exclusive events.

c. Calculate probabilities using addition and multiplication rules.

d. Define independence and calculate probability of independent events.

e. Calculate probabilities using combinatorics, such as combinations and permutations.

f. Define and calculate conditional probabilities.

g. State Bayes's theorem and use it to calculate conditional probabilities.

1.1.1 Basic Concepts in Probability

1.1.1.1 Probability Spaces and Events

Sample point and sample space: A **sample point** is the simple outcome of a random experiment. The probability space (also called sample space) is the collection of all possible sample points related to a specified experiment. When the experiment is performed, one of the sample points will be the outcome. The probability space is the "full set" of possible outcomes of the experiment.

Mutually exclusive outcomes: Outcomes are **mutually exclusive** if they cannot occur simultaneously. They are also referred to as **disjoint** outcomes.

Exhaustive outcomes: Outcomes are **exhaustive** if they combine to be the entire probability space, or equivalently if at least one of the outcomes must occur whenever the experiment is performed.

Event: Any collection of sample points, or any subset of the probability space is referred to as an **event**. We say that the "event A has occurred" if the experimental outcome was one of the sample points in A.

Union **of events A and B**: $A \cup B$ denotes the union of events A and B, and consists of all sample points that are in either A or B (or both).

Union of events $A_1, A_2, \cdots, A_n$: $A_1 \cup A_2 \cup \cdots \bigcup A_n = \bigcup_{i=1}^{n} A_i$ denotes the **union of the events** $A_1, A_2, \cdots, A_n$, and consists of all sample points that are in at least one of the A_i's. This definition can be extended to the union of infinitely many events.

Intersection of events $A_1, A_2, \cdots, A_n$: $A_1 \cap A_2 \cap \cdots \cap A_n = \bigcap_{i=1}^{n} A_i$ denotes the **intersection of the events** $A_1, A_2, \cdots, A_n$, and consists of all sample points that are simultaneously in all A_i's. ($A \cap B$ may also be denoted $A \cdot B$ or AB).

Mutually exclusive events $A_1, A_2, \cdots, A_n$: Two events are **mutually exclusive** if they have no sample points in common, or equivalently, if they have **empty intersection**. Events $A_1, A_2, \cdots, A_n$ are mutually exclusive if $A_i \cap A_j = \emptyset$ for all $i \neq j$, where $\emptyset$ denotes the empty set with no sample points. Mutually exclusive events cannot occur simultaneously.

Exhaustive events $B_1, B_2, \cdots, B_n$: If $B_1 \cup B_2 \cup \cdots \cup B_n = S$, the entire probability space, then the events $B_1, B_2, \cdots, B_n$ are referred to as **exhaustive events**.

Complement of event A: The **complement of event** A consists of all sample points in the probability space that are **not in A**. The complement is denoted A' or $\bar{A}$ or A^c and is equal to $\{x : x \notin A\}$. When the underlying random experiment is performed, to say that the complement of A has occurred is the same as saying that A has not occurred.

Subevent (or subset) A of event B: If event B contains all the sample points in event A, then A is a **subevent** of B, denoted $A \subset B$. The occurrence of the event A implies that the event B must also have occurred.

Partition of event A: Events $C_1, C_2, \cdots, C_n$ form a **partition of event A** if i) $A = \bigcup_{i=1}^{n} C_i$ and ii) the C_i's are mutually exclusive.

DeMorgan's Laws:

1. $(A \cup B)' = A' \cap B'$; to say that $A \cup B$ has not occurred is to say that A has not occurred and B has not occurred; this rule generalizes to any number of events;

$$\left(\bigcup_{i=1}^{n} A_i\right)' = (A_1 \cup A_2 \cup \cdots \cup A_n)' = A_1' \cap A_2' \cap \cdots \cap A_n' = \bigcap_{i=1}^{n} A_i'.$$

2. $(A \cap B)' = A' \cup B'$, to say that $A \cap B$ has not occurred is to say that either A has not occurred or B has not occurred (or both have not occurred); this rule generalizes to any number of events,

$$\left(\bigcap_{i=1}^{n} A_i\right)' = (A_1 \cap A_2 \cap \cdots \cap A_n)' = A_1' \cup A_2' \cup \cdots \cup A_n' = \bigcup_{i=1}^{n} A_i'.$$

Indicator function for event A: The function

$$I_A(x) = \begin{cases} 1 & \text{if } x \in A \\ 0 & \text{if } x \notin A \end{cases}$$

is the **indicator function** for event A, where x denotes a sample point. $I_A(x)$ is 1 if event A has occurred.

Module 2

Financial Mathematics

Financial Mathematics (or Interest Theory, as some like to call it) is a fundamental subject for all actuaries that applies concepts and mathematical tools to calculate values for various cash flow streams for different financial tasks, e.g., loan and bond valuation, asset/liability management, and investment income analysis. Financial Mathematics has a specialized language, and you will need to spend adequate time to fully understand the meanings of all the concepts and terms introduced in this module. Most problems in this module will be word problems (rather than just formulas), and it is very difficult to solve these problems unless you understand the language being used.

Over the years, most actuarial students have found that the best way to prepare for actuarial exams is to work a very large number of problems (hundreds and hundreds of problems). You should plan to spend a significant proportion of your study time working problems and reviewing solutions to gain a deeper understanding of the concepts and tools of financial mathematics.

In this module, we will review all key components of financial mathematics as an integral part of your actuarial skill set. You can expect to apply interest theory regularly throughout your career. Mastering the topics covered in this module will provide you with not only a solid foundation for the ACIA Capstone Exam, but also a valuable toolbox for understanding financial and economic matters on and off the job.

This page is intentionally left blank.

Contents

2.1 Time Value of Money

Interest theory deals with the **time value of money**. For example, a dollar invested today at 6% **interest** per year will be worth \$1.06 one year from today. In this example, the dollar that is invested today is called the **principal**, and the \$0.06 increase in value is called **interest**.

Because a dollar invested today can grow to *more* than one dollar a year from now, receiving a dollar today has *greater value* than receiving a dollar one year from now. In other words, money has a "time value". Consistent with the study outcomes listed in the CIA education syllabus, we aim to provide a thorough discussion of the terms and concepts needed to assess the value of a payment, considering both the *amount* of the payment and its *timing*.

*****FROM THE CIA EDUCATION SYLLABUS*****

A2.1 Time value of money

a. Define and recognize the definitions of the following terms: interest rate; simple interest; compound interest; accumulation function; future value; current value, present value, and net present value; discount factor; discount rate; convertible monthly; nominal rate; effective rate; inflation and real rate of interest; force of interest; and equation of value.

b. Given any three of the interest rate, period, present value, and future value, calculate the remaining item based on simple or compound interest.

c. Solve time value of money equations involving variable force of interest.

d. Given any one of the effective interest rate, the nominal interest rate convertible monthly, the effective discount rate, the nominal discount rate convertible monthly, or the force of interest, calculate all of the other items.

e. Write the equation of value given a set of cash flows and an interest rate.

2.1.1 Present Value and Future Value

The value of an investment today (time 0) is its **present value** (PV); its value n periods from today is called its **future value** (FV) as of time n. More broadly, if we know the value of an investment as of a particular date and we want to find its value as of an *earlier* date, we are calculating a *present value* as of the earlier date. And if we want to find the value as of a *later* date, then we are calculating a *future value* (or an **accumulated value**) as of that later date. If funds are invested at a compound interest rate of i per period for n periods, the basic relationships are:

$$FV = PV\,(1+i)^n \qquad\qquad PV = \frac{FV}{(1+i)^n} \tag{2.1.1}$$

Example 2.1.1.

Let $n = 10$ and $i = 0.06$ (an interest rate of 6% per year, compounded annually).

a) If $PV = 1,000$, then $FV = 1,000 \times 1.06^{10} = 1,790.85$.

b) If $FV = 1,000$, then $PV = \frac{1,000}{1.06^{10}} = 558.39$.

Calculation a) demonstrates that if we invest 1,000 today at 6% interest, in ten years it will have accumulated to a future value of 1,790.85.

Calculation b) shows that if we need 1,000 ten years from now, we can accumulate that amount by investing 558.39 now at 6% interest.

Calculator Note

The BA II Plus calculator has 5 "**Time Value of Money**" keys:

- [N] Number of periods
- [I/Y] Interest rate per period (usually per year)
- [PV] Present Value
- [PMT] Periodic Payment
- [FV] Future Value

These keys are used for performing calculations in the **Time Value of Money (TVM) Worksheet**.

In this subsection, we will not look at any problems that involve periodic payments. The [PMT] key will be used later. Using the other four keys, we can solve compound interest problems like Example **2.1.1**, as we illustrate next.

To begin any new problem, it is wise to clear the Time Value of Money (TVM) registers to erase any entries from prior problems. Note that the legend "CLR TVM" appears above the [FV] key on the BA II Plus calculator. To clear the TVM registers use the keystrokes [2nd] [CLR TVM]. This sets all 5 of the TVM values to 0.

Module 4

Portfolio Management and Corporate Finance

This module provides important background knowledge on Portfolio Management and Corporate Finance, key areas for understanding how investment strategies and corporate financial decisions are made. Although this material will not be directly tested on the ACIA Capstone Exam, it serves as an essential context for understanding broader actuarial applications. Concepts such as portfolio optimization, risk management, and corporate financing strategies are crucial to making informed financial decisions in real-world scenarios.

In this module, we will cover important financial topics like asset pricing models. These models are used to assess the expected return of an asset given its risk and form the basis for understanding how investments are priced. In the next module about *Option Pricing*, some of the principles become crucial when determining the value of financial derivatives.

This page is intentionally left blank.

Contents

4.1 Mean-Variance Portfolio Theory

*****FROM THE CIA EDUCATION SYLLABUS*****

A3.4 Mean-Variance Portfolio Theory

a. Estimate the risk and return of an asset, given appropriate inputs.

b. Calculate the risk and expected return of a portfolio of many risky assets, given the expected return, volatility, and correlation of returns of the individual assets.

c. Explain the assumptions of mean-variance theory and understand the importance of the mean/standard-deviation diagram and the resulting efficient market frontier.

d. Calculate the optimal portfolio, locate the capital market line, and describe the limitations of this approach.

e. Describe how portfolio risk can be reduced through diversification across multiple securities or across multiple asset classes.

4.1.1 Risk and Return

2365This module is about portfolio mathematics. You will be introduced to the classical model for computing the required return on an asset / a portfolio / a real project given a specified level of risk. We will define how risk can be measured and discuss some simple ideas in asset allocation.

The first section is about how one can convert asset prices to returns and calculate summary statistics from a series of returns. We assume basic knowledge of probability and statistics, and this subsection is a review of them put in the context of portfolios.

4.1.1.1 Calculating Historical Returns and Volatility

One-period Realized Return

Suppose that you are a shareholder (i.e., you own shares of a certain firm). Over a certain period of time, you have **capital gain** or capital loss due to changes in share price:

$$\text{Capital gain} = \text{change in share price} \times \text{number of shares owned.}$$

For example, if you own two hundred shares of a stock whose current price is \$35.2 per share, and after one day the stock price jumps to \$36.4, your capital gain is $(36.4 - 35.2) \times 200 = \$240$. If the share price drops so that the change in the share price is negative, the capital gain would also be negative. This means that you suffer a capital loss. In the following, we will refer to capital losses as negative capital gains.

You may wonder why the capital gain is really a "gain." If you do not sell the stock, then you cannot capitalize the gain of \$240 because you do not receive any cash. But since you have the right to sell the stock and realize that \$240, we can still regard the capital gain as a gain.

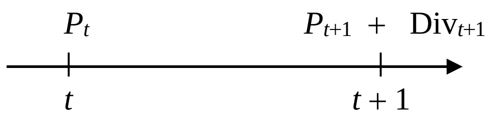
Sometimes, when the company is profitable, it may pay *dividends* to its shareholders. Therefore, when you calculate your **total gain** / loss, you must include dividends:

$$\text{Total dollar return} = \text{Dividend income} + \text{Capital gain}.$$

For example, if the stock pays a dividend of \$1 per share after one day, then the total dollar return is

$$1 \times 200 + 240 = \$440.$$

The realized return is the rate of return that actually occurs over a particular time period. Let P_t and P_{t+1} be the price per share of the stock at time t and $t+1$, respectively, and Div_{t+1} be the dividend paid at time $t+1$ per share:



Then the **realized return** on the stock over $(t, t+1]$ is defined as

$$\text{Realized return} = \text{Dividend yield} + \text{Capital gain rate}$$

$$R_{t+1} = \frac{\text{Div}_{t+1}}{P_t} + \frac{P_{t+1} - P_t}{P_t}$$

In our example, the percentage change in share price is $(36.4 - 35.2)/35.2 = 3.41\%$, and the dividend yield is $1/35.2 = 2.84\%$, giving a realized return of 6.25%.

The concept of calculating realized return applies to all financial instruments, including bonds and preferred stocks. Any intermediate cash flows paid during a year can be treated as dividends.

Example 4.1.1.

Consider a 20-year corporate bond with a par value of 1,000. The bond, which pays no coupons, was issued one year ago with a price of \$140.50. The current effective annual market interest rate is 11%. Compute the realized return on the bond over the last year.

Solution. The price of the zero-coupon bond is now

$$1,000 \times 1.11^{19} = \$137.6776,$$

giving a capital gain of $137.6776 - 140.50 = \$2.82236$. Since the zero-coupon bond does not pay an intermediate cash flow during the previous year, the realized return is

$$-2.82236/140.5 \times 100\% = -2.009\%.$$

□

Module 5

Finance Common and Specialized: Introductory Derivatives and Option Pricing

Module 5 delves into one of the most dynamic areas of actuarial finance: derivatives and option pricing. Derivatives are financial contracts that derive their value from underlying assets, such as stocks, bonds, or interest rates. Understanding derivatives and how they are priced is crucial for actuaries, as these instruments are used extensively in risk management, insurance, and investment strategies.

One of the most well-known types of derivatives is the option, which gives the holder the right (but not the obligation) to buy or sell an asset at a specified price before or at a certain date. For example, a call option on a stock allows an investor to buy that stock at a fixed price, no matter how high the stock's price rises in the future. This concept may seem straightforward, but determining the fair value of options is where things get complex and fascinating. This module covers both **discrete** (Binomial Model) and **continuous** (Black-Scholes Model) methods of option pricing, providing you with the mathematical tools to evaluate financial risks and develop effective hedging strategies.

This module is built on the foundations of probability and finance covered in earlier modules. Mastering these concepts will not only prepare you for the ACIA Capstone Exam but also equip you with valuable skills for managing financial risks in real-world scenarios, from hedging strategies in investment portfolios to pricing guarantees in life insurance products.

This page is intentionally left blank.

Contents

5.1 Introductory Derivatives – Forwards and Futures

*****FROM THE CIA EDUCATION SYLLABUS*****

A3.9 Introductory Derivatives – Forwards and Futures

a. Distinguish between long and short positions for both assets (including short selling of stocks) and derivatives on assets.

b. Recognize the transaction costs affecting profit calculations for both assets and derivatives on assets (including commissions and bid-ask spread).

c. Recognize the definitions of the following terms relating to both forward contracts and prepaid forward contracts: forward contract, prepaid forward contracts, outright purchase, fully leveraged purchase, payoff of long and short forward, net profit of long and short forward.

d. Determine payoffs and profits for both long and short positions on forward contracts.

e. Calculate prices for both forward contracts and prepaid forward contracts on stocks with no dividends, continuous dividends, and discrete dividends.

f. Construct a synthetic forward from the underlying stock and a risk-free asset, and identify arbitrage opportunities when the synthetic forward price is different from the market forward price.

g. Recognize the definitions of the following terms: marking to market, margin balance, maintenance margin, and margin call.

h. Evaluate an investor's margin balance based on changes in asset values.

5.1.1 Stock as an Underlying Asset

There are thousands of financial instruments in today's financial world. In this manual, you will be introduced an important class of financial instruments known as derivatives. As its name suggests, derivatives are "derived" from some more fundamental financial instruments known as underlying assets. Before we start our journey of derivatives, we need to understand the underlying assets. In this subsection we focus on stocks.

5.1.1.1 Financial Markets

This part provides with you some factual information. The chance that you would be tested on these materials is slim.

You must have heard of the term "**financial market**". In economics,

- a market refers to a variety of systems, institutions, procedures and also the possible buyers and sellers of a certain good or service;
- a financial market is a market in which people trade different kinds of financial securities, including **bonds**.

Shopping in a financial market is quite different from shopping in a mall. When you enter a grocery store, you become a potential buyer. The store, which is the seller, lists the prices of the goods. You pay the price, and then you get the goods. Sometimes you may ask for the goods to be delivered to you within say 10 days after purchase if the goods is too bulky or is stored in a warehouse. You can also be a seller, too, if you open your own grocery store, in which case you may be setting prices. Notice that in any transaction, there would be two parties: the buyer and the seller. Later on we will use "**long**" to refer to buyers and "**short**" to refer to sellers.

Now let us consider what would happen if you want to trade (which can mean "buy" or "sell") stocks, bonds, or derivatives. Such financial assets are certainly not traded in a grocery. As a matter of fact, many financial assets do not physically exist; they only exist on electronic records and represent an ownership or right to do something. Nowadays, you would not get a large pile of bond certificates and coupons when you buy a coupon bond! The trading would typically involve at least 4 steps:

1. The buyer and seller locate each other and agree on a price.
2. The trade is then **cleared**. It means that the obligations of the buyer and the seller are specified. For example, the buyer agrees to pay the seller by a specified date and the seller agrees to deliver the asset upon receiving payment.
3. The trade is then **settled**. It means that the buyer and the seller fulfill the obligations.
4. A change of ownership of the financial asset is recorded. The trade is completed.

In real life, it is hard for buyers and sellers to find each other. Step 1 is facilitated by brokers, dealers and sometimes exchanges. Stocks are usually traded in an organized exchange, where rigorous rules that govern trading and information flows exist. In the US, we have, for example, the New York Stock Exchange (NYSE) for stocks and the Chicago Board Options Exchange (CBOE) for many derivatives. For other assets you can go to an **over-the-counter (OTC) market**. For OTC markets there is not a physical location where trading takes place. Trading is also less formal. In both cases the buyer or seller would contact a broker, who then contacts a market maker to create a trade.

> **Market makers** are traders who will buy assets from customers who wish to sell, and sell financial assets to customers who wish to buy.

Module 6

Long-Term Actuarial Mathematics Common

The Long-Term Actuarial Mathematics (LTAM) Common Module is a crucial element of the ACIA Capstone Exam, concentrating on the mathematical frameworks used to assess risk and uncertainty in long-term insurance products, including life insurance and annuities. This module builds on the foundational concepts established in previous modules, exploring the fundamental topics of LTAM including long-term insurance coverage, survival models and estimation, present value random variables, and long-term insurance premium calculation. Understanding these topics is vital, as they provide the tools needed to determine the price of life insurance-related products and evaluate the long-term financial obligations of insurers, where liabilities can extend over many years or even decades.

The primary focus of this module is exam preparation. It emphasizes the importance of developing a systematic approach to problem solving. Practicing a wide variety of problems enhances speed and accuracy and helps you identify common patterns and strategies. Re-visiting previously solved exercises reinforces your understanding, allowing you to confidently tackle exam questions. This module also offers insights into interpreting exam language, enabling you to discern what is required under exam conditions quickly.

To maximize your learning experience in this module, it is essential to have a solid foundation in probability theory and financial mathematics. The module will guide you through both theoretical and practical dimensions of long-term insurance mathematics, equipping you for both exam success and a rewarding career in the actuarial field.

This page is intentionally left blank.

Contents

6.1 Long-Term Insurance Coverages

*****FROM THE CIA EDUCATION SYLLABUS*****

A 4.1 Long-term Insurance Coverages

a. Describe the long-term coverages in insurance (life, health, and general), annuities, and retirement benefits (e.g., pensions, retiree health care, etc.).

b. Describe the similarities and differences among the long-term coverages identified in a).

c. Describe the appropriate models to be used to calculate expected present values, premiums or contributions, and reserves for each long-term coverage.

This subsection serves as an introduction to the coverages. It contains descriptions of various life and health insurance products and pension plans. Let us start with reviewing three definitions:

Insurable interest: When you have an insurable interest in an entity, it means you would suffer a financial loss if that entity is damaged. It is a basic requirement in law that a person paying life insurance premium has insurable interest in the insured, or else the insurance would have significant *moral hazard*: the policyholder may wish the insured to die. For example, you cannot take out a life insurance on a university professor and expect to collect a death benefit when he or she dies. Here are a few examples of insurable interest: family members (spouse, children), business partner, employees.

Group and individual insurance: Group insurance is an insurance that covers a group of people. An employer can purchase life and health coverages for its employees on a yearly basis. Individual insurances are insurances that one can purchase for just him/herself (or for his or her family). Group insurances help reduce *adverse selection* (that is, the risk that the insurance products attract high risk individuals, leading to excessive claims) and this allows insurance companies to offer lower rates.

Assessmentism: This means the premiums each year are set to match the cost of claim each year. This is commonly seen in group life and health insurances. Individual life insurances typically have level premiums because assessmentism would mean increasing premium as the probability of death increases with age. This discourages policy renewal, and if the insureds became very ill, they would price themselves out of the insurance right as they needed it most.

6.1.1 Traditional Life Insurance Contracts

Term insurance

- It can be used to protect a family (policyholder's spouse and children) in the event of the death of the policyholder.

- It can also be used to protect businesses against losses arising from the deaths of key employees (key person insurance / company owned life insurance).
- Most have a level sum insured, but some companies offer decreasing term insurances which can be used to pay the outstanding loan balance of a mortgage.
- *Convertibility*: some term insurances can be converted to whole life or endowment when it ends without further underwriting.
- *Renewable term*: allows the policyholder to renew the current contract when it ends without further underwriting (with an increased premium).

Whole life insurance

- Older lives can use it to cover funeral expenses or to reduce inheritance taxes, younger lives can use it as a passive investment vehicle with substantial death benefit.
- For regular premium contract, typically the premium is payable to some maximum age (80 – 90).
- The insurance company assumes the premium received are invested in very safe financial vehicles so that the investment return is low, and this results in an interest spread (difference between returns of a typical investment and the return on the premium) which is used to cover the company's profit and margin for adverse experience.
- *Cash value*: Lapse occurs if the policyholder fails to pay a premium, in such case the policyholder would be eligible for a cash value or surrender value that represents the investment part of the paid premiums (which can be very low or even 0 if lapse occurs in early years).
- *Lapse-supported insurance*: For whole life insurance designed for older lives, the cash value distributed to policyholders is close to zero. The insurance company makes profit for a lapsed policy and this profit can reduce the premiums of the pool.
- *Stranger owned life insurance (STOLI)*: In some jurisdictions, it is legal for a policyholder to sell his or her policy to a third party for a one-time cash payment. After the sale, the purchaser would continue to make payments and become the policy's beneficiary. Such arrangements can be illegal in many countries because the beneficiary has no insurable interest in the life of the insured. For STOLI to make sense to both parties, the price must be greater than the cash value and less than the death benefit.

Participating insurance

- It is also known as par insurance in North America or with-profit insurance outside North America.
- Some of the profits earned on the invested premiums are shared with the policyholder so that the contract looks more attractive.
- Losses on the invested premiums are not shared.
- The profit share can take different forms, for example, cash refunds and reduced premiums (in the form of dividends) which are common in North America, or increased sum insured (in the form of bonuses) which are commonly seen in the UK or in Australia.

Module 7

Long-Term Actuarial Mathematics Specialized

The Long-Term Actuarial Mathematics (LTAM) Specialized module emphasizes the actuarial techniques required for effective management of long-term insurance reserves and pension plans. In this module, you will learn to calculate and interpret various reserve types, including net premium, modified, gross premium, and expense reserves, as well as their associated probabilities, means, and variances. Key profit measuressuch as expected profit, actual profit, and internal rate of returnwill also be explored, along with the application of approximation methods like the uniform distribution of deaths and Euler methods.

Additionally, the module covers the intricacies of pension plans, contrasting defined benefit and defined contribution plans, and addressing retiree health-care considerations. You will gain proficiency in modeling common states and decrements, utilizing parametric and Markov chain models, to evaluate defined benefit plans based on participant data and valuation assumptions. Calculating actuarial accrued liabilities, normal costs, and understanding the impact of changes in underlying assumptions such as mortality and salary increases will further enhance your analytical skills.

By mastering these advanced topics, you will be well-prepared to tackle real-world challenges in long-term actuarial practice, ensuring a solid foundation for both exam success and a flourishing career in the field.

This page is intentionally left blank.

Contents

7.1 Long-Term Insurance Reserves

*****FROM THE CIA EDUCATION SYLLABUS*****

A 4.5 Long-term Insurance Reserves

a. Calculate and interpret the following reserve types: net premium, modified, gross premium, expense.

b. Calculate and interpret probabilities, means, variances, and percentiles of random variables associated with these reserves, including future-loss random variables.

c. Calculate and interpret common profit measures such as expected profit, actual profit, gain, gain by source and period, internal rate of return, profit margin, and break-even year.

d. Apply appropriate approximation methods such as uniform distribution of deaths, constant force, Woolhouse, and Euler.

7.1.1 Net Premium Policy Values

We have discussed the equivalence principle, which states that the mean of the loss-at-issue random variable ${}_0L$ is zero, or

$$\text{APV, at time 0, of future benefit} = \text{APV, at time 0, of all future net premiums.}$$

After a period of time, however, there will no longer be equivalence. Consider, for example, a 20-payment year whole life insurance of 1,000 on (40). At the time of issuance, the APV of the future benefit is $1{,}000A_{40}$, while the APV of the 20 yearly payments is $\pi\ddot{a}_{40:\overline{20|}}$. The value of π is set to $\dfrac{1{,}000A_{40}}{\ddot{a}_{40:\overline{20|}}}$ such that the two APVs are equal. However, after 20 years, if the policyholder still survives, then the APV of future benefit is $1{,}000A_{60}$, while the APV of future premiums is 0. So

$$\text{APV, at time 20, of future benefit} > \text{APV, at time 20, of all future net premiums.}$$

Clearly, the insurer has a financial obligation while the policyholder does not have any in this example. The **policy value** (or reserve) for a policy is a measure of the value of such financial obligation of the insurer.

There are many ways to calculate policy values. In this subsection, we discuss **net premium policy value** (or benefit reserve and net premium reserve in older exams). The net premium policy value at time h is the conditional mean of the difference between the present value of future benefits and the present value of future net premiums, the conditioning event being the survivorship of the policyholder to time h.

7.1.1.1 The Prospective Approach

Net Premium Policy Values for a General Insurance Policy

Let us introduce the prospective loss random variable ${}_hL$, which is defined as follows:

$$\begin{aligned} {}_hL = & \text{ Present value, at time } h\text{, of all future benefits} \\ & - \text{Present value, at time } h\text{, of all } \textbf{net} \text{ premiums.} \end{aligned}$$

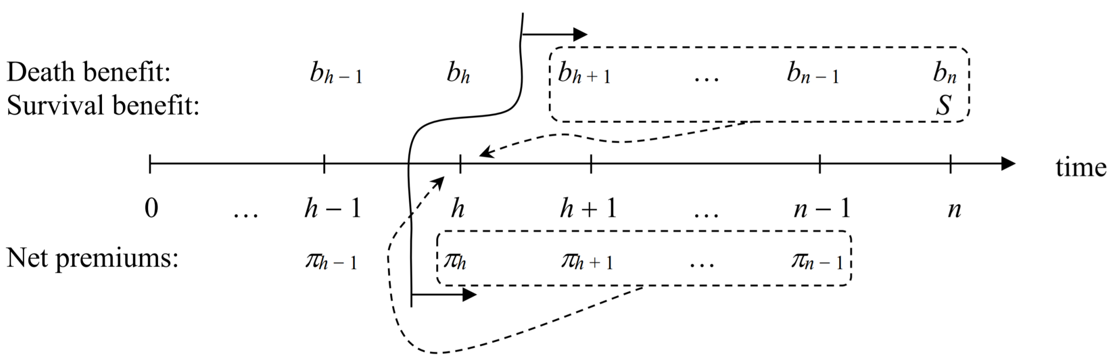
${}_hL$ is defined conditionally on the policy to be in force at time h, that is, $T_x > h$ (or $K_x \geq h$ for discrete policies). The net premium policy value, at time h, for a policy on (x), is defined as

$$ {}_hV = \mathrm{E}({}_hL). $$

To fix ideas let's look at a general n-year fully discrete insurance on (x) in which

1. Net premiums are payable annually at the beginning of each policy year, the payment to be made at time j being π_j, $j = 0, 1, \ldots, n-1$;
2. The death benefit is payable at the end of the policy year of death, the payment to be made if death occurs during the j-th policy year is b_j, $j = 1, 2, \ldots, n$;
3. The survival benefit, payable at time n if the policyholder survives to time n, is S.

If the policyholder survives to time h, then the future benefits and premiums are those on the right-hand side of the 'cut':



The prospective loss random variable ${}_hL$ (for $h = 0, 1, 2, \ldots, n$) is

$$ {}_hL = \begin{cases} b_{h+K_{x+h}+1} v^{K_{x+h}+1} - \displaystyle\sum_{j=0}^{K_{x+h}} \pi_{h+j} v^j & K_{x+h} < n-h \\ S v^{n-h} - \displaystyle\sum_{j=0}^{n-h-1} \pi_{h+j} v^j & K_{x+h} \geq n-h. \end{cases} $$

The probability distribution of K_{x+h} is

$$ \Pr(K_{x+h} = j) = {}_{j|}q_{x+h}, \ \Pr(K_{x+h} \geq j) = {}_jp_{x+h}. $$

As a result, the net premium policy value at the end of year h is

$$ {}_hV = \mathrm{E}({}_hL) = \sum_{j=0}^{n-h-1} b_{h+j+1} v^{j+1} \ {}_{j|}q_{x+h} + S \ {}_{n-h}E_{x+h} - \sum_{j=0}^{n-h-1} \pi_{h+j} v^j \ {}_jp_{x+h}. $$

Module 8

Short-Term Actuarial Mathematics Common

The Short-Term Actuarial Mathematics (STAM) module is a critical part of the ACIA Capstone Exam, focusing on the practical applications of probability theory to model risk and uncertainty in short-term insurance coverages. It builds on the probability foundations covered in the previous module and expands into specific areas such as loss models, frequency and severity distributions, credibility theory, and pricing of insurance products. These concepts are vital for understanding the mechanics behind short-term insurance, including health, auto, and property insurance, where the timing and amount of future claims are uncertain.

Due to the time constraint on the exam, a crucial aspect of exam preparation is the ability to work quickly and efficiently. Working through numerous problems and examples is one of the best ways to build speed and confidence. By revisiting problems you have already solved, you reinforce your understanding and become more adept at recognizing patterns and question types, which is essential for quick problem identification on the exam. This module emphasizes systematic approaches and shortcuts wherever possible, aiming to equip you with techniques that streamline your problem-solving process. You will also find comments on the interpretation of the language of the exam questions, helping you better understand what is being asked under pressure.

The primary focus of this module is exam preparation, but from time to time you will encounter commentary on the underlying theory behind specific concepts. These insights provide a deeper understanding, helping you not just pass the exam but also grasp how these principles apply in real-world actuarial work. For instance, understanding the nuances of loss models and how they apply to pricing and reserving is not just a test skill-it is a fundamental aspect of actuarial practice that influences decision-making in insurance companies.

In order to get the most out of this module, it is important that you already have a solid grasp of the basics covered in the probability theory module. You should also be comfortable with calculus, particularly differential and integral calculus, as these skills are essential for solving many of the problems in this module. As we move forward, this module will guide you through both the theoretical and practical aspects of short-term insurance mathematics, preparing you for both the exam and your future career as an actuary.

This page is intentionally left blank.

Contents

This page is intentionally left blank.

8.1 Severity Models

*****FROM THE CIA EDUCATION SYLLABUS*****

A4.7 Severity Models

a. Calculate the basic distributional quantities: moments, percentiles, generating functions.

b. Describe how changes in parameters affect the distribution.

c. Recognize classes of distributions, including extreme value distributions, suitable for modeling the distribution of severity of loss and their relationships.

d. Apply the following techniques for creating new families of distributions: multiplication by a constant, raising to a power, exponentiation, and mixing.

e. Identify the applications in which each distribution is used and the reasons why.

f. Apply the distribution to an application given the parameters.

g. Compare two distributions based on various characteristics of their tails, including moments, ratios of moments, limiting tail behavior, hazard rate function, and mean excess function.

8.1.1 Parametric Distributions

In this subsection, we explore a range of parametric distribution families commonly used to model claim severities.

8.1.1.1 Introduction

For any random variable, the mean and variance, skewness, etc., are “parameters” of the distribution (some of these might be infinite) that can be calculated if the form of the distribution is known, or they can be estimated when a sample of data is available. There is also the notion of a **parametric distribution**, which means that the random variable X has a PDF, PF or CDF which is formulated in terms of parameters. In this definition, “parametric distribution” refers to a collection of distributions based on the set of all possible values of the parameters. Some examples are as follows.

1. The **uniform distribution** on the interval $[0, \theta]$ has PDF:

$$f(x) = \begin{cases} 1/\theta & \text{if } 0 \le x \le \theta, \\ 0 & \text{otherwise.} \end{cases}$$

This is a parametric distribution with parameter θ. The mean is $\theta/2$, and other distribution quantities such as variance, skewness, etc. are formulated in terms of the parameter θ.

2. The **normal distribution** with parameters μ (mean) and σ (standard deviation) has PDF:

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-(x-\mu)^2/2\sigma^2}, \ -\infty < x < \infty.$$

3. The **Poisson distribution** with mean (parameter) $\lambda > 0$ is a discrete, non-negative integer-valued random variable with PF

$$f(x) = \mathbb{P}(X = x) = \frac{e^{-\lambda}\lambda^x}{x!}, x = 0, 1, \cdots.$$

4. The **exponential distribution** with mean (parameter) $\theta > 0$ has PDF

$$f(x) = \frac{1}{\theta} e^{-x/\theta}, \ x > 0,$$

and CDF:

$$F(x) = 1 - e^{-x/\theta}.$$

8.1.1.2 Scale Distributions and Scale Parameters

A continuous parametric distribution is a **scale distribution** if cX is a member of the set of distributions whenever X is a member and $c > 0$ is a constant. θ is called a **scale parameter** if the corresponding parameter for the distribution of cX is $c\theta$. It is possible for a distribution to be a scale distribution and yet not have a scale parameter (the lognormal distribution is an example).

Example 8.1.1.

The exponential distribution has a PDF of the form:

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

or equivalently, CDF of the form

$$F(x) = 1 - e^{-x/\theta},$$

where $\theta > 0$.

Show that this is a scale distribution and that θ is the scale parameter.

Solution. If $Y = cX$ and $c > 0$, then

$$\mathbb{P}(Y \le y) = \mathbb{P}\left(X \le \frac{y}{c}\right) = 1 - e^{-(y/c)/\theta} = 1 - e^{-y/c\theta}.$$

Y has the CDF of an exponential distribution with parameter $c\theta$. Therefore the exponential family is a scale family and the exponential parameter θ is a scale parameter. □

Module 9

Short-Term Actuarial Mathematics Specialized

Module 9 focuses on the principles and methods of ratemaking and reserving for short-term insurance coverages, a crucial area of actuarial practice. This module corresponds to the **Exam 2b Specialized Short-term (SST)** section of the CIA Capstone Exam, focusing on the specialized skills needed to effectively price and reserve short-term insurance products.

Within this module, you will cover rating factors, exposure, and experience rating, all of which are essential for aligning premium calculations with policyholders' risk profiles. Key techniques like the chain ladder, average cost per claim, and Bornhuetter-Ferguson methods are introduced to estimate unpaid losses from run-off triangles, a standard method for projecting future liabilities. Additionally, you will learn premium calculation methods, including pure premium and loss ratio calculations, based on historical claims experience.

This module builds on the probability and statistical foundations from earlier modules, guiding you in applying these concepts to the practical tasks of ratemaking and reserving. By mastering these topics, you will be prepared to address both exam questions in the SST specialization and practical actuarial challenges in short-term insurance.

This page is intentionally left blank.

Contents

This page is intentionally left blank.

9.1 Pricing and Reserving for Short-Term Insurance Coverages

*****FROM THE CIA EDUCATION SYLLABUS*****

A4.15 Pricing and Reserving for Short-Term Insurance Coverages

a. Explain the role of rating factors and exposure.

b. Describe the different forms of experience rating.

c. Describe and apply techniques for estimating unpaid losses from a run-off triangle using the chain ladder, average cost per claim, and Bornhuetter-Ferguson methods.

d. Describe the underlying statistical models for the methods in c).

e. Calculate premiums using the pure premium and loss ratio methods.

9.1.1 Short-Term Insurance Loss Reserving

9.1.1.1 Basic Elements of Loss Reserving

Here, we review some basic factors used in in loss reserving. Some of them are variations on concepts that have been introduced earlier in this manual.

- Average claim frequency:

$$f = \frac{\text{number of incurred claims}}{\text{units of earned exposure}}$$

- Average loss severity:

$$S = \frac{\text{dollars of incurred losses}}{\text{number of incurred claims}} = \text{ average payment per claim}$$

- Aggregate loss distribution: combines frequency and severity.
- Pure premium: the expected loss per unit of exposure.
- Loss cost:

$$\text{Loss cost } = f \cdot S = \frac{\text{dollars of incurred losses}}{\text{units of earned exposure}}$$

- An interest rate and an estimated payout pattern must be assumed.

Individual Claim File Estimates

An individual claim file will contain information on loss severity, time to settlement and final payment, and an inflation assumption. The case reserves for a particular year and line of business is the aggregate of individual claim files.

9.1.1.2 Case Reserves Plus and Expected Loss Ratio Methods

Case Reserves Plus

The **case reserve** is the aggregate of individual claim file estimates, which is also incurred but not paid loss amounts. The following amounts combined are called the **bulk reserve** or **gross IBNR reserve**.

- Pure IBNR (Incurred But Not Reported).
- RBNR (Reported But Not Recorded).
- Future adjustments of case reserves on known claims.
- Files that are closed but may reopen.

The "**case reserves plus**" method combines case reserves and IBNR reserves.

Expected Loss Ratio Method For a particular block of business and policy period, we have the following definitions:

$$\text{Estimated Ultimate Losses}_{LR} = \text{Expected Loss Ratio} \times \text{Earned Premium},$$
$$\text{Estimated Loss Reserve}_{LR} = \text{Estimated Ultimate Losses}_{LR} - \text{Losses Paid To Date},$$
$$\text{Total Estimated Loss Reserve}_{LR} = \sum \text{Estimated Loss Reserve},$$

where the sum is taken over all blocks of business and policy years.

This may be the only method appropriate for new lines of business. Also, there may be mandatory regulatory minimum **loss ratios** that must be applied. This method requires that an expected loss ratio be chosen and the earned premium calculated. The standard definition of earned premium is the amount of total premiums collected by an insurance company over a period that have been earned based on the ratio of the time passed on the policies to their effective life. This pro-rated amount of "paid in advance" premiums have been earned and now belong to the insurer.

9.1.1.3 Chain-Ladder (Loss Development Triangle) Method

This method is described in the following illustration.

Suppose that we have actual claims paid data for each year from 2010 up to and including 2017. For each **claim occurrence year** or **accident year (AY)** from 2010 to 2017 the claim loss settlement amount is known for that year (Column 0) and for each following year up to and including 2017. This is summarized in Table 9.1 below.

"**Development Year**" refers to the number of years after AY. For instance, across the row for AY 2015, column 0 (development year 0 is calendar AY) is 3408 (340,800), the claims paid in 2015, column 1 is 3168 (316,800), the claims paid in 2016 for accidents that occurred in 2015 (development year 1 is calendar year AY+1, which is 2016 if AY = 2015) and column 2 is 2325 (232,500), the claims paid in 2017 for accidents that occurred in 2015 (development year 2 is calendar year AY+2, which is 2017 for AY = 2015). There are no entries to the right of the diagonal in the table because those claims paid are for later years (2018, ...) and data is only available to the end of 2017 in this example. The diagonal entries for each row (each AY) are the 2017 entry of claims (in 1,000's) paid for that particular AY. "Incremental" refers to year-by-year payment.

Module 10

Statistics Learning I

The Statistical Learning module is a critical part of the ACIA Capstone Exam. You will learn the general tools available for constructing and evaluating predictive models (e.g., training/test set split, cross-validation), and the technical details of specific types of models and techniques (e.g., linear models, generalized linear models, regression-based time series models, decision trees, principal components analysis, and clustering).

This is the first module on the topic of statistical learning, of which the primary goal is to introduce the fundamental concepts of statistical learning and discuss the most common techniques, aka, the (generalized) linear regression model. To ensure learning effectiveness, the module contains both comprehensive introductions on the conceptual materials, rigorous mathematical derivations for the theory of models and extensive practical implementation examples for hand-on exposures.

To fully benefit from this module, it is crucial to have a solid foundation in prior probability modules and statistics lectures in your university study. As you progress, this module will equip you with the knowledge and skills necessary to succeed in the exam and advance your actuarial career.

This page is intentionally left blank.

Contents

10.1 Basic of Statistical Learning

*****FROM THE CIA EDUCATION SYLLABUS*****

A 5.1 Basics of Statistical Learning

a. Explain the types of modeling problems and methods, including supervised versus unsupervised learning and regression versus classification.

b. Explain the common methods of assessing model accuracy.

c. Employ basic methods of exploratory data analysis, including data checking and validation.

10.1.1 A Primer on Statistical Learning

To set the scene for the whole module, we start with a broad overview of *statistical learning*, which is a big umbrella encompassing linear regression and many other statistical techniques. We first describe the general problem setting and terminology that pervade statistical learning. Then we present some key theoretical concepts and practical issues that commonly arise in this subject and introduce some quantitative tools to address these issues and evaluate the performance of a statistical learning method.

It is important to note that the concepts and tools introduced in this section are universally applicable in the sense that they apply to essentially all types of models, and will be illustrated in the context of any specific types of models (linear models, GLMs, and decision trees, in particular) in this and later modules.

10.1.1.1 Fundamental Concepts

What is statistical learning? In a typical framework where there is an output of interest, we have a set of input variables that may provide valuable information for predicting or interpreting the output. This "input-output" structure is a hallmark of statistical learning, and our task is to build a model that identifies the (potentially complex) relationships between the inputs and the output.

To fix ideas, let y be, as usual, the response variable and $\mathbf{x} = (x_1, \ldots, x_p)$ be the vector of accompanying predictors. We are interested in relationships between the (quantitative) response and the predictors in the general form of

$$\boxed{\underbrace{y}_{\text{response}} = \underbrace{f(\mathbf{x})}_{\substack{\text{systematic part}\\ \text{(signal)}}} + \underbrace{\varepsilon}_{\substack{\text{idiosyncratic part}\\ \text{(noise)}}},} \tag{10.1.1}$$

where:

- f is a fixed (non-random) but unknown real function connecting the predictors and the response variable. This function carries the systematic information that the predictors offer about the response variable. Different types of statistical learning methods (or models) are distinguished by the structural form of this function. For linear regression models, which are a prominent and perhaps the simplest example of a statistical learning method, the function $f(\mathbf{x}) = \beta_0 + \beta_1 x_1 + \cdots + \beta_p x_p$ is linear in the regression coefficients $\beta_0, \beta_1, \ldots, \beta_p$.

- ε is a random error term carrying information that cannot be captured by the systematic component of the model. Statistically speaking, it is independent of $\mathbf{x}$ and has a zero mean.

Although (10.1.1) looks abstract and the exam will probably not test it directly, it will provide a useful framework for thinking about statistical learning. For convenience, we will refer to f and ε respectively as the *signal function* and the *noise*, which are widely used terms originally stemming from engineering. We are interested in the signal, but the data we have is "contaminated" with noise. The goal of statistical learning is to filter out the noise and use a variety of "statistical" tools and techniques to "learn" as much about the signal f as possible from the data, e.g., the sign and magnitude of the coefficients $\beta_0, \beta_1, \ldots, \beta_p$ for linear models. This knowledge about the signal can then serve as the basis for making inferences and predictions.

Two main objectives of statistical learning. As we have just discussed, the main objectives of statistical learning are two-fold:

1. *Inference:* One objective of statistical learning is to use the estimated signal function, $\hat{f}$, to understand the association between the response y and the predictors $x_1, \ldots, x_p$. We can acquire such an understanding by looking at the form of $\hat{f}$, which can reveal:

   - Which predictors are strongly associated with y (those represented in $\hat{f}$)

   - How the important predictors are related to y (is the relationship positive or negative, linear or non-linear?)

2. *Prediction:* Based on $\hat{f}$, another important objective of statistical learning is to make accurate predictions for the response variable on sets of predictor values of interest, say $\mathbf{x}_*$. To do so, we simply evaluate $\hat{f}$ at $\mathbf{x}_*$ and produce a prediction using the equation

   $$\hat{y}_* = \hat{f}(\mathbf{x}_*),$$

   where both $\hat{f}$ and $\mathbf{x}_*$ are known.

   The error of $\hat{y}_*$ as a predictor of the true (random) target $y_* = f(\mathbf{x}_*) + \varepsilon_*$ can be broken down into two components:

   - The *reducible error* stemming from the quality of $\hat{f}$ as an estimate of f. This kind of error is reducible because it can be potentially "reduced" with the use of more appropriate statistical learning techniques (or by increasing the sample size). The focus of statistical learning lies in the development of analytic techniques intended to minimize the reducible error.

Module 11

Statistics Learning II

In this module, we extensively discuss Extended Linear Models, linear Mixed Models and Time-Series Models. You will need to understand the theoretical fundamentals covered in this module, including the purposes, motivations, mechanics, advantages, limitations, best practices, and distinctions between various techniques, to answer questions in the Capstone Exam. Additionally, you will need time to gain experience in applying these concepts comprehensively to real-life business problems in this field.

The primary goal of this study module is to help you develop both a conceptual understanding of and hands-on experience with these topics as effectively and efficiently as possible. To achieve this, the module is divided into three key sections:

- **Extended Linear Models:** This part introduces the fundamentals of extended linear models, including assumptions behind different forms of the model, model diagnosis, variable selections and shrinkage methods.
- **Linear Mixed Models:** This part covers the general theoretical foundations of LMM, its implementations with different types of data and the practical implementation examples for each scenario.
- **Time-Series Models:** This section focuses on the fundamentals of time series analysis, introducing the key components of a typical time series and forecasting techniques.

To fully benefit from this module, it is crucial to have a solid foundation in the basics covered in prior statistical learning modules. As you progress, this module will equip you with the knowledge and skills necessary to succeed in the exam and advance your actuarial career.

This page is intentionally left blank.

Contents

11.1 Extended Linear Models

*****FROM THE CIA EDUCATION SYLLABUS*****

A 5.3 Extended Linear Modles

a. Understand the assumptions behind different forms of the extended linear model and be able to select the appropriate model, such as ordinary least squares, generalized linear model, ANOVA, generalized additive models, local regression, lasso ridge regression, partial least squares, and principal component analysis (PCA) regression.

b. Evaluate models developed using the extended linear model approach.

c. Understand the algorithms behind the numerical solutions for the different forms of the extended linear model family to enable the interpretation of output from the statistical software employed in modeling and to make appropriate modeling choices when selecting modeling options.

d. Understand and be able to select the appropriate model structure for an extended linear model given the behavior of the dataset to be modeled.

e. Identify the advantages and limitations of modeling techniques.

In this section, we will explore several extensions of the linear models discussed in the previous module in greater detail. To ensure completeness, we will also revisit the foundational aspects of linear models, including model setup, parameter estimation, hypothesis testing, and diagnostic methods for identifying potential anomalies.

11.1.1 Linear Models

11.1.1.1 Model Formulation and Parameter Estimation

Model formulation. In linear models, it is postulated that the response variable Y is related to p predictors $X_1, X_2, \ldots, X_p$, whose values are given or known (i.e., non-random), via the approximately[i] linear relationship

$$Y = \underbrace{\beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_p X_p}_{\text{regression line}} + \epsilon, \tag{11.1.1}$$

[i]The linear relationship is only approximate due to the presence of the random error ϵ.

where:

- p is the number of predictors.
- $\beta_0, \beta_1, \ldots, \beta_p$ are unknown regression coefficients (or parameters).
- ϵ is the (unobservable) *random error term* that accounts for the fluctuation of Y about its mean.

A statistical model of the form (11.1.1) is called a *general linear model* or *multiple linear regression* (MLR) *model*, with Y "regressed" on $X_1, \ldots, X_p$. For convenience, we refer to

β_0 as the *intercept term* (which is the expected value of Y when all X_j's are zero) and

β_j as the *regression coefficient* (or *regression parameter*, *slope parameter*) attached to the jth predictor for $j = 1, \ldots, p$. Since $\beta_j = \partial \mathrm{E}[y]/\partial x_j$, we can interpret β_j as the expected change in y (also called the expected effect on y) per unit change in X_j, holding all other predictors fixed.[ii]

Almost always, we assume that ϵ is normally distributed. The normality assumption will be central to much of the statistical inference and effectively implies that the response Y is also normally distributed with

$$Y \sim \mathrm{N}(\beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_p X_p, \sigma^2).$$

In particular, the mean of Y, which is

$$\mathrm{E}[Y] = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_p X_p,$$

is linear in $\beta_0, \beta_1, \ldots, \beta_p$. In general, a linear model entails the mean of the response variable being linear in the regression parameters β_j's, but not necessarily in the predictors. For example, the quadratic model

$$\mathrm{E}[Y] = \beta_0 + \beta_1 X + \beta_2 X^2$$

is also a linear model, although $\mathrm{E}[Y]$ is non-linear in X.

Data structure. To estimate the unknown parameters $\beta_0, \beta_1, \ldots, \beta_p$ and σ^2, we assume that we are given n independent pairs of observations $\{(Y_i, \mathbf{X}_i)\}_{i=1}^n = \{(Y_i, X_{i1}, X_{i2}, \ldots, X_{ip})\}_{i=1}^n$ governed by

$$Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \cdots + \beta_p X_{ip} + \epsilon_i, \quad i = 1, \ldots, n, \tag{11.1.2}$$

where $\epsilon_i \overset{\text{i.i.d.}}{\sim} \mathrm{N}(0, \sigma^2)$ for some common variance σ^2. In spreadsheet form, the data structure is:

Observation	**Response Variable**	**Explanatory Variables**			
	Y	X_1	X_2	$\cdots$	X_p
1	Y_1	X_{11}	X_{12}	$\cdots$	X_{1p}
2	Y_2	X_{21}	X_{22}	$\cdots$	X_{2p}
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\ddots$	$\vdots$
n	Y_n	X_{n1}	X_{n2}	$\cdots$	X_{np}

[ii]The "ceteris paribus" condition is part of the definition when computing partial derivatives as you learn in a multi-variable calculus class.

Module 12

Statistics Learning III

In this module, we extensively discuss the following remaining topics regarding Statistical Learning: Principal Components Analysis(PCA), Decision Trees and Cluster Analysis.

Each of these methods is specifically designed to address challenges associated with high-dimensional data: Principal Components Analysis (PCA) is a widely used unsupervised learning technique that reduces the dimensionality of a dataset by transforming it into a smaller set of variables while preserving most of the original information. Decision trees partition the high-dimensional predictor space into non-overlapping regions, helping to segment data into homogeneous groups that facilitate analysis and prediction. Cluster Analysis organizes observations into distinct, non-overlapping subgroups, revealing hidden patterns in the data.

To develop a comprehensive understanding of these techniques, you must grasp the theoretical fundamentals covered in this module and focus on connecting and distinguishing them from the methods learned in previous modules. The knowledge gained here will be essential for answering questions in the Capstone Exam and solving real-world business problems.

A solid foundation in the concepts from prior statistical learning modules is crucial to fully benefit from this material. As you progress, this module will equip you with the knowledge and skills necessary to excel in the exam and advance your actuarial career.

This page is intentionally left blank.

Contents

This page is intentionally left blank.

12.1 Principal Components Analysis

*****FROM THE CIA EDUCATION SYLLABUS*****

A 5.7 Principal Component Analysis

a. Define principal components.

b. Interpret the results of a principal component analysis, considering loading factors and proportion of variance explained.

c. Explain uses of principal components.

d. Understand and apply principal component analysis.

12.1.1 PCA: Fundamental Ideas

Motivation: What does PCA do? When you do predictive modeling in practice and deal with real datasets, you will often see tens if not hundreds of variables, many of which are correlated in subtle ways, which can make the visualization and analysis of the dataset a very challenging task. For such a high-dimensional dataset, examining the pairwise relationships among the large number of variables is not only a daunting task—if there are $p = 10$ (a reasonably small number) features, then there are already a total of $\binom{p}{2} = \binom{10}{2} = 45$ pairs to examine!—but also unlikely to show us the "big picture" that lies behind the data. We need alternative data exploration tools that are more effective and efficient.

PCA is a commonly used unsupervised learning technique that transforms a high-dimensional dataset into a smaller, much more manageable set of variables that capture most of the information in the original dataset (hence the qualifier "principal,"[i] meaning major or representative). Known as the *principal components* (PCs), these composite variables are linear combinations of the existing variables generated to collectively simplify the dataset, reducing its dimension and making it more amenable to data exploration and visualization. This technique especially lends itself to *highly correlated data*, for which a few PCs are enough to represent most of the information in the full dataset.

What exactly are the PCs? Throughout this section, suppose for concreteness that we are given n observations, each containing the measurements of p features,[ii] $X_1, X_2, \ldots, X_p$. (There is no response variable Y in unsupervised learning.)

[i] PCA is commonly misspelled as "principle components analysis." Do not make this mistake!

[ii] We prefer the terms "variables" and "features" rather than "predictors" in this subsection because there is nothing to predict in an unsupervised learning setting.

The data matrix is the $n \times p$ matrix given by[iii]

$$\mathbf{X} = (x_{ij})_{\substack{i=1,\dots,n \\ j=1,\dots,p}} = \begin{array}{c} \\ 1 \\ 2 \\ \vdots \\ n \end{array} \overset{\begin{array}{cccc} 1 & 2 & \cdots & p \end{array}}{\begin{pmatrix} x_{11} & x_{12} & \cdots & x_{1p} \\ x_{21} & x_{22} & \cdots & x_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ x_{n1} & x_{n2} & \cdots & x_{np} \end{pmatrix}}.$$

Here x_{ij} is the value of the jth variable for the ith observation. For each x_{ij}:

- i is the index for the observations (the rows), ranging from 1 to n.
- j is the index for the variables or features (the columns), ranging from 1 to p.

Without loss of generality, we assume that the observations of each feature have been centered to have a zero mean, i.e., the sample mean of the jth feature is $\bar{x}_j = \sum_{i=1}^n x_{ij}/n = 0$ for any $j = 1, 2, \dots, p$ (the column means of $\mathbf{X}$ are zero). If this is not true, we can consider the mean-centered values $x'_{ij} = x_{ij} - \bar{x}_j$ and perform PCA on these zero-mean feature values. We will explain why this zero-mean assumption does not change anything very shortly.

Mathematically, the PCs are composite variables defined as *normalized linear combinations* of the original set of features. For $m = 1, 2, \dots, M$, where $M\ (\leq p)$ is the number of PCs to use, the mth PC is given by

$$Z_m = \sum_{j=1}^p \phi_{jm} X_j = \underbrace{\phi_{1m}X_1 + \phi_{2m}X_2 + \cdots + \phi_{pm}X_p}_{\text{linear combinations of } X_1, X_2, \dots, X_p}, \tag{12.1.1}$$

where the coefficients $\phi_{1m}, \phi_{2m}, \dots, \phi_{pm}$ corresponding to $X_1, X_2, \dots, X_p$ are the *loadings* (a.k.a. *weights*) of the mth PC. In (12.1.1), you can see that we are compressing the p features into a single metric via a linear combination, which has the advantage of quickly summarizing the data so long as the PC loadings are well-chosen, and we will see very shortly how to determine the loadings for each PC. For now, keep in mind that:

- The PCs are constructed from the features, so the sum in (12.1.1) is taken over j (from 1 to p), which is the index for the features, but not i.
- The generic symbol ϕ_{jm} for PC loadings is indexed as follows:

$$\begin{aligned} j &: \text{ Index for the variables (features)} \\ m &: \text{ Index for PC} \end{aligned}$$

Given the data matrix $\mathbf{X}$ and the PC loadings, we can compute the realized values of Z_m by

$$\boxed{z_{im} = \phi_{1m}x_{i1} + \phi_{2m}x_{i2} + \cdots + \phi_{pm}x_{ip} = \sum_{j=1}^p \phi_{jm}x_{ij}, \quad i = 1, \dots, n,} \tag{12.1.2}$$

[iii] We are not fitting GLMs, so there is no need to create a column of 1's representing the intercept, unlike a design matrix.

Module 13

Predictive Analytics

The Predictive Analytics (PA) module is a critical part of the ACIA Capstone Exam, focusing on the practical applications of the knowledge of probability, mathematical statistics, and selected analytical techniques to to a business problem and see first hand how things play out. You will need to understand the underlying theory, including the purposes, motivations, mechanics, advantages, limitations, best practices, and distinctions between various predictive analytic techniques. Additionally, you will need time to gain hands-on experience in fitting and interpreting predictive models in R, as well as practice in effectively communicating your insights in writing. The written-answer format will test the material in greater breadth and depth, and assess your higher-level thinking, e.g., can you describe a certain concept or explain why something is true? You have to know how things work, at least at a conceptual level, and organize your thoughts in words.

The primary goal of this study module is to help you develop both a conceptual understanding of and hands-on experience with predictive analytics as effectively and efficiently as possible. To achieve this, the module is divided into three key sections:

- **Crash Course in R:** This part introduces the fundamentals of R programming, including basic concepts and data visualization techniques.
- **Conceptual Background of Predictive Analytics:** This part covers the theoretical foundations of predictive analytics, supplemented with practical implementation examples.
- **Communication Skills in Predictive Analytics:** This part focuses on clear and audience-appropriate communication, guiding you to structure reports effectively and present complex ideas in straightforward language.

To fully benefit from this module, it is crucial to have a solid foundation in the basics covered in prior statistical learning modules. As you progress, this module will equip you with the knowledge and skills necessary to succeed in the exam and advance your actuarial career.

This page is intentionally left blank.

Contents

13.1 A Primer on Predictive Analytics

*****FROM THE CIA EDUCATION SYLLABUS*****

A 5.10 Predictive Analytics Problem Definition and Tools

a. Understand the different types of predictive modeling problems.

b. Write and execute basic commands in R using RStudio or other coding languages.

c. Translate a vague question into one that can be analyzed with statistics and predictive analytics to solve a business problem.

d. Consider factors such as available data and technology, significance of business impact, and implementation challenges to define the problem.

This section streamlines the scattered material and presents a coherent introduction to predictive analytics. The fundamental concepts introduced in this section are universally applicable in the sense that they apply to essentially all types of models.

13.1.1 Basic Terminology

Predictive analytics in a nutshell. It would be helpful to introduce the three main categories of predictive modeling problems to answer the question of "What exactly is predictive analytics":

- *Descriptive:* Descriptive analytics focuses on what happened in the *past* and aims to "describe" or explain the observed trends by identifying the relationships between different variables in the data.

  Example. If you saw an increase in the lapse rate among the policyholders of a certain line of business, what kind of policyholders had the highest tendency to lapse? This is a question addressed by descriptive analytics.

- *Predictive:* Predictive analytics focuses on what will happen in the *future* and is concerned with making accurate "predictions."

  Example. For a prospective policyholder with certain characteristics, what is their predicted probability of lapse? The ability to make such a prediction will be useful for identifying future policyholders who will have a lower probability of lapse and contribute to the profitability of an insurer.

- *Prescriptive:* Prescriptive analytics uses a combination of optimization and simulation to investigate and quantify the impact of different "prescribed" actions in different scenarios.

  Example. If we reduce the premium by a certain amount, how will this affect the lapse rate?

Not surprisingly, We are predominantly focusing on *predictive analytics* in this section, although the three modeling approaches are often mutually complementary, not contradictory. (In fact, the actuarial exams even say that "it is common to use the term predictive analytics to refer to all three cases.") Throughout this exam, we will be using a vast set of statistical tools for understanding and "predicting" a variable of interest based on a set of closely related variables. Using predictive analytics, we can provide a data-driven response to many questions of practical interest. Here are some motivating examples:

- *(COVID-19)* Every day many patients all over the world go to hospitals to get tested for COVID-19. How do medical professionals classify them into "positive" and "negative" based on the wealth of biomedical measurements collected, for example, through a nasal swab (presence of antibodies, antigens, etc.)? How is the classification rule formulated? We all know how important it is to make accurate classifications, but it is especially important to identify infected patients so that they can be treated as soon as possible and isolated.
- *(Salary)* What role do factors such as age, gender, education level, number of exams passed, and years of experience play in determining the salary of actuaries? We all know that the association between the number of exams passed and salary is strongly positive, but a predictive model can quantify this association precisely.
- *(Email spam)* We all are inundated with junk emails every day. How do email service providers detect spam? Are there keywords that are indicative of spam? While we want junk emails to be filtered, of course we don't want important emails (e.g., those from your boss) to end up in the junk mailbox. Unlike the COVID-19 example above, the classification rule has to be sensitive, but specific.
- *(Ratemaking)* Predictive analytics also plays a pivotal role in insurance *ratemaking*, a topic that deals with setting premiums, or rates, based on the loss experience and risk characteristics of policyholders. With the aid of predictive models, an insurance company will have a scientific basis for calibrating premiums for different types of policyholders in a way that ensures the financial viability of the company.

Although the examples above involve entirely different contexts, all of them have something in common. There is always an output (or outcome) of interest, which can be numeric (salary, premium, injury rates) or categorical (positive/negative, email/spam), and we have at our disposal a collection of input variables that may offer potentially useful information for predicting or understanding the output. This "input-output" setting is characteristic of predictive analytics in general, and our job is to develop a model teasing out the (possibly complex, overlapping) contributions of the inputs to the output.

Classification of variables. Predictive analytics requires data, often with a large number of observations and variables. Generally speaking, there are two ways to classify variables in a predictive analytic context: By their role in the study (intended use) or by their nature (characteristics).

- *By role:* The variable that we are interested in predicting is called the *target variable*, or simply the *target* (a.k.a. ***response variable***, *dependent variable*, *output variable*, *outcome variable*). Despite the target variable being our interest, in most situations, we cannot change the target directly, but we have control over some associated variables that can be used to predict the target. These variables go by different names, such as *predictors*, ***explanatory variables***, ACTEX Learning

Who We Are

A Benefit Corporation Experienced at Teaching Actuaries!

Experienced

More than 50 years experience helping students prepare and pass actuarial exams! We are an eLearning technology and education company leveraging experts in the field to constantly update our learning content in a format that works for you.

Trustworthy

ACTEX Learning is a leading US based provider of study materials for actuarial exams. Our authors and content contributors are renowned academics and Actuaries that are proud to have their names on the cover of our manuals and textbooks!

Mission Focused

We are a Benefit Corporation focusing on the mission of accessible high quality actuarial education. We're dedicated to empowering actuarial students by offering test prep materials that are not just effective and efficient but also tailored to suit every type of student.

Become an ACTEX Champion

Join our Global ACTEX Champion Program and bring the benefits to your Actuarial Club. To learn more about the program, scan the QR code on the right. If you have any questions or would like to speak to a Champion Coordinator, please do not hesitate to reach out to us at champions@actexlearning.com.

ACTEX Learning

Use GOAL to Practice What You've Learned

- Over 22,000 exam-style problems with detailed solutions
- 3 learning modes (Practice, Quiz, Simulated Exams)
- 3 levels of difficulty (Core, Advanced and Mastery)
- You control your topics and sub-topics
- Dedicated instructor support

GOAL is currently available for the following SOA & CAS Exams:

Exam P
1,050+ Questions

Exam FM
1,500+ Questions

Exam FAM-L
1,300+ Questions

Exam FAM-S
900+ Questions

Exam FAM
2,300+ Questions

Exam ALTAM
1,400+ Questions

Exam ASTAM
1,300+ Questions

Exam SRM
1,150+ Questions

Exam MAS-I
1,050+ Questions

Exam MAS-II
850+ Questions

Exam CAS 5C
500+ Questions

Exam CAS 6US
550+ Questions

Exam CAS 6CA
650+ Questions

Use GOAL Score to Gauge Your Exam Readiness

Measure how prepared you are to pass your exam with a tool that suits any study approach. A GOAL Score of 70 or above indicates readiness.

Your score is broken into categories, allowing you to study efficiently by concentrating on problem areas. GOAL Score quantifies your exam readiness by measuring both your performance and the consistency of your performance. Your GOAL Score also analyzes your strengths and weaknesses by category, topic, and level of difficulty.

Scan to Learn More