Institute of Actuaries of India

Subject
CS2 - Risk Modelling and Survival Analysis
Core Principles

For 2024 Examinations

CS2 - Risk Modelling and Survival Analysis

Syllabus for the 2024 Examinations

This syllabus includes information to support the study of this subject. It will guide you through what you need to learn, application of learning as well as the skills that you need to develop. Information regarding the assessment of this subject is also included.

This syllabus includes:

- Aim of the subject
- How this subject links across the Qualifications
- Subject topics and topic weightings
- Subject objectives
- Assessment information

Aim

The aim of the Risk Modelling and Survival Analysis Core Principles subject is to provide a grounding in mathematical and statistical modelling techniques that are of relevance to actuarial work, including stochastic processes and survival models and their application.

Link Across the Qualifications

Associateship Qualification

This subject assumes that the student is competent with the material covered in CS1 and the required knowledge for that subject. CM1 and CM2 apply the material in this subject to actuarial and financial modelling.

Fellowship Qualification

Topics in this subject are further built upon in SP1, SP7, SP8 and SP9.

Topics & Topic Weighting

1. Random variables and distributions for risk modelling [20%]
2. Time series [20%]
3. Stochastic processes [25%]
4. Survival models [25%]
5. Machine Learning [10%]

Objectives

1 Random variables and distributions for risk modelling [20%]

Statistical distributions suitable for modelling the variables and risks that arise within insurance contracts.

1.1 Loss distributions, with and without risk sharing

1.1.1 Properties of the statistical distributions that are suitable for modelling individual and aggregatelosses

1.1.2 Concepts of excesses (deductibles) and retention limits

1.1.3 Operation of simple forms of proportional and excess of loss reinsurance

1.1.4 Calculate the distribution and corresponding moments of the claim amounts paid by the insurerand the reinsurer in the presence of excesses (deductibles) and reinsurance

1.1.5 Estimate the parameters of a failure time or loss distribution when the data is complete, or whenit is incomplete, using maximum likelihood and the method of moments

1.1.6 Fit a statistical distribution to a data set and calculate appropriate goodness-of-fit measures

1.2 Compound distributions and their applications in risk modelling

1.2.1 Construct models appropriate for short-term insurance contracts in terms of the numbers of claims and the amounts of individual claims

1.2.2 Major simplifying assumptions underlying the models in 1.2.1

1.2.3 Compound Poisson distribution and apply the result that the sum of independent random variables, each having a compound Poisson distribution, also has a compound Poisson distribution

1.2.4 Mean, variance and coefficient of skewness for compound binomial, Compound Poisson and compound negative binomial random variables

1.2.5 Loss distributions for both the insurer and the reinsurer after the operation of simple forms of proportional and excess of loss reinsurance where underlying losses take the forms given in 1.2.4

1.3 Introduction to copulas

1.3.1 Characterise a copula as a multivariate distribution function that is a function of the marginal distribution functions of its variates, and explain how this allows the marginal distributions to beinvestigated separately from the dependency between them

1.3.2 Meaning of the terms ‘dependence or concordance’, ‘upper and lower tail dependence’, and state in general terms how tail dependence can be used to help select a copula suitable for modelling particular types of risk

1.3.3 Know the form and characteristics of the Gaussian copula and the Archimedean family of copulas

1.4 Introduction to extreme value theory

1.4.1 Recognise extreme value distributions, suitable for modelling the distribution of severity of lossand their relationships

1.4.2 Calculate various measures of tail weight and interpret the results to compare the tail weights

2 Time series [20%]

Statistical concepts for modelling, fitting and forecasting data that is indexed by time.

2.1 Understand the core concepts underlying time series models

2.1.1 General properties of stationary, I(0), and integrated, I(1), univariate time series

2.1.2 Stationary random series

2.1.3 Stationary random series with a filter applied

2.1.4 Know the notation for backwards shift operator, backwards difference operator and the conceptof roots of the characteristic equation of time series

2.1.5 Concepts and basic properties of Autoregressive (AR), Moving Average (MA), Autoregressive Moving Average (ARMA) and Autoregressive Integrated Moving Average (ARIMA) time series

2.1.6 Concept and properties of discrete random walks and random walks with normally distributedincrements, both with and without drift

2.1.7 Basic concept of a multivariate autoregressive model

2.1.8 Cointegrated time series

2.1.9 Univariate time series models have the Markov property and how to rearrange a univariatetime series model as a multivariate Markov model

2.2 Applications of time series models

- **2.2.1** The processes of identification, estimation and diagnosis of a time series, the criteria for choosing between models and the diagnostic tests that may be applied to the residuals of a time series after estimation
- **2.2.2** Other non-stationary, non-linear time series models
- **2.2.3** Simple applications of a time series model, including random walk, autoregressive and cointegrated models, as applied to security prices and other economic variables
- **2.2.4** Develop deterministic forecasts from time series data, using simple extrapolation and moving- average models, applying smoothing techniques and seasonal adjustment when appropriate

3 Stochastic processes [25%]

Application of Markov models to model time-indexed risk and claim data arising primarily in insurance andother appropriate business-related scenarios.

3.1 Stochastic processes

- **3.1.1** Stochastic processes, in particular a counting process

3.1.2 A stochastic process according to whether it:

- operates in continuous or discrete time
- has a continuous or a discrete state space
- is a mixed type

- **3.1.3** Applications of mixed processes
- **3.1.4** Markov property in the context of a stochastic process and in terms of filtrations

3.2 Understand and apply a Markov chain

- **3.2.1** Essential features of a Markov chain model
- **3.2.2** Calculate the stationary distribution for a Markov chain in simple cases
- **3.2.3** Understand and apply systems of frequency-based experience rating in terms of a Markovchain
- **3.2.4** Time-inhomogeneous Markov chain model and describe simple applications
- **3.2.5** Markov chains can be used as a tool for modelling and how they can be simulated

3.3 Define and apply a Markov process

- **3.3.1** Essential features of a Markov process model
- **3.3.2** Poisson process, derive the distribution of the number of events in a given time interval, derivethe distribution of inter-event times and apply these results
- **3.3.3** Kolmogorov equations for a Markov process with time-independent and time/age-dependenttransition intensities
- **3.3.4** Kolmogorov equations in simple cases
- **3.3.5** Simple survival models, sickness models and marriage models in terms of Markov processes and describe other simple applications
- **3.3.6** Kolmogorov equations for a model where the transition intensities depend not only on age/time,but also on the duration of stay in one or more states
- **3.3.7** Sickness and marriage models in terms of duration-dependent Markov processes and describe other simple applications
- **3.3.8** The Markov jump processes and how it can be used as a tool for modelling and how they canbe simulated

4 Survival models [25%]

Description, estimation and use of statistical models for the time until an event occurs.

4.1 Concepts of survival models

4.1.1 Model of lifetime or failure time from age x as a random variable

4.1.2 Consistency condition between the random variable representing lifetimes from different ages

4.1.3 Distribution and density functions of the random future lifetime, the survival function, the force ofmortality or hazard rate, and derive relationships between them

4.1.4 Understand the actuarial symbols tpx and tqx and derive integral formulae for them

4.1.5 Gompertz and Makeham laws of mortality

4.1.6 Curtate future lifetime from age x and state its probability function

4.1.7 Understand the symbols ex and °ex and derive an approximate relation between them. Define the expected value and variance of the complete and curtate future lifetimes and deriveexpressions for them

4.1.8 Two-state model of a single decrement and compare its assumptions with those of the random lifetime model

4.2 Understand the estimation procedures for lifetime distributions

4.2.1 Various ways in which lifetime data may be censored

4.2.2 Estimation of the empirical survival function in the absence of censoring and what problems areintroduced by censoring

4.2.3 Kaplan–Meier (or product limit) estimator of the survival function in the presence of censoring,compute it from typical data and estimate its variance

4.2.4 Nelson–Aalen estimator of the cumulative hazard rate in the presence of censoring, compute itfrom typical data and estimate its variance

4.2.5 Models for proportional hazards and how these models can be used to estimate the impact ofcovariates on the hazard

4.2.6 Cox model for proportional hazards, derive the partial likelihood estimate in the absence of ties,and state the asymptotic distribution of the partial likelihood estimator

4.3 Derive maximum likelihood estimators for transition intensities

4.3.1 Identify an observational plan in respect of a finite number of individuals observed during a finiteperiod of time, and define the resulting statistics, including the waiting times

4.3.2 Understand the likelihood function for constant transition intensities in a Markov model oftransfers between states given the statistics in 4.3.1

4.3.3 Identify maximum likelihood estimators for the transition intensities in 4.3.2 and state theirasymptotic joint distribution

4.4 State the Poisson approximation to the estimator in 4.3.3 in the case of a single decrement

4.5 Transition intensities dependent on age (exact or census)

4.5.1 Dividing the data into homogeneous classes, including subdivision by age and sex

4.5.2 The principle of correspondence and its fundamental importance in the estimation procedure.

4.5.3 Specify the data needed for the exact calculation of a central exposed to risk (waiting time)depending on age and sex

4.5.4 Calculate a central exposed to risk given the data in 4.4.3

4.5.5 Understand how to obtain estimates of transition probabilities

4.5.6 Identify the assumptions underlying the census approximation of waiting times.

4.5.7 Concept of the rate interval

4.6 Graduation and graduation tests

- **4.6.1** Statistical tests of the comparison of crude estimates with a standard mortality table testing for:
    - the overall fit
    - the presence of consistent bias
    - the presence of individual ages where the fit is poor
    - the consistency of the 'shape' of the crude estimates and the standard table. For each test,describe:
    - the formulation of the hypothesis
    - the test statistic
    - the distribution of the test statistic using approximations where appropriate
    - the application of the test statistic
- **4.6.2** Reasons for graduating crude estimates of transition intensities or probabilities, and state thedesirable properties of a set of graduated estimates
- **4.6.3** How to test for smoothness of a set of graduated estimates
- **4.6.4** The process of graduation by the following methods, and the advantages and disadvantages ofeach (the candidate will not be required to carry out a graduation):
    - Parametric formula
    - Standard table
    - Spline functions.
- **4.6.5** How should the tests in 4.5.1 be amended to compare crude and graduated sets of estimates
- **4.6.6** How should the tests in 4.5.1 be amended to allow for the presence of duplicate policies
- **4.6.7** Carry out a comparison of a set of crude estimates and a standard table or of a set of crudeestimates and a set of graduated estimates

4.7 Mortality projection

- **4.7.1** Approaches to forecasting of future mortality rates based on extrapolation, explanation andexpectation, as well as their advantages and disadvantages
- **4.7.2** Lee-Carter, age-period-cohort and p-spline regression models for forecasting mortality.
- **4.7.3** Use an appropriate computer package to apply the models in 4.6.2 to a suitable mortality dataset
- **4.7.4** Identify main sources of error in mortality forecasts

5 Machine learning [10%]

5.1 Elementary principles of machine learning.

- **5.1.1** Bias/variance trade-off and relationships with model complexity
- **5.1.2** Cross-validation to evaluate models on unseen data, and estimate hyper-parameters
- **5.1.3** Understand how regularisation can be used to reduce overfitting in highly parameterised models.
- **5.1.4** The use of software to apply supervised learning techniques, to solve regression andclassification problems.
- **5.1.5** the use of metrics such as precision, recall, F1 score and diagnostics such as the ROC curve and confusion matrix to evaluate the performance of a binary classifier.
- **5.1.6** Unsupervised learning techniques (principal component analysis, K-means clustering) toreduce data dimensionality, identify latent substructure and detect anomalies.

Assessment

Combination of a one-hour and forty-five-minutes computer-based data analysis and statistical modelling assignment and a three- hours and fifteen-minutes written examination.

END