Institute
and Faculty
of Actuaries

CS2 – Risk Modelling and Survival Analysis Core Principles

Syllabus

for the 2023 exams

June 2022

CS2 – Risk Modelling and Survival Analysis Core Principles

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 particular relevance to actuarial work, including stochastic processes and survival models and their application.

Competences

On successful completion of this subject, a candidate will be able to:

1 describe and use statistical distributions for risk modelling.

2 describe and apply the main concepts underlying the analysis of time series models.

3 describe and apply Markov chains and processes.

4 describe and apply techniques of survival analysis.

5 describe and apply basic principles of machine learning.

Links to other subjects

This subject assumes that the candidate 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.

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

Syllabus topics

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%)

These weightings are indicative of the approximate balance of the assessment of this subject between the main syllabus topics, averaged over a number of examination sessions.

The weightings also have a correspondence with the amount of learning material underlying each syllabus topic. However, this will also reflect aspects such as:

- the relative complexity of each topic and hence the amount of explanation and support required for it.
- the need to provide thorough foundation understanding on which to build the other objectives.
- the extent of prior knowledge that is expected.
- the degree to which each topic area is more knowledge- or application-based.

Skill levels

The use of a specific command verb within a syllabus objective does not indicate that this is the only form of question that can be asked on the topic covered by that objective. The Examiners may ask a question on any syllabus topic using any of the agreed command verbs, as are defined in the document 'Command verbs used in the Associate and Fellowship examinations'.

Questions may be set at any skill level: Knowledge (demonstration of a detailed knowledge and understanding of the topic), Application (demonstration of an ability to apply the principles underlying the topic within a given context) and Higher Order (demonstration of an ability to perform deeper analysis and assessment of situations, including forming judgements, taking into account different points of view, comparing and contrasting situations, suggesting possible solutions and actions and making recommendations).

In the CS subjects, the approximate split of assessment across the three skill types is 20% Knowledge, 65% Application and 15% Higher Order skills.

Detailed syllabus objectives

1 Random variables and distributions for risk modelling (20%)

1.1 Loss distributions, with and without risk sharing.

1.1.1 Describe the properties of the statistical distributions that are suitable for modelling individual and aggregate losses.

1.1.2 Explain the concepts of excesses (deductibles) and retention limits.

1.1.3 Describe the operation of simple forms of proportional and excess of loss reinsurance.

1.1.4 Derive the distribution and corresponding moments of the claim amounts paid by the insurer and 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 when it 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 Describe the major simplifying assumptions underlying the models in 1.2.1.

1.2.3 Define a compound Poisson distribution and show that the sum of independent random variables, each having a compound Poisson distribution, also has a compound Poisson distribution.

1.2.4 Derive the mean, variance and coefficient of skewness for compound binomial, compound Poisson and compound negative binomial random variables.

1.2.5 Repeat 1.2.4 for both the insurer and the reinsurer after the operation of simple forms of proportional and excess of loss reinsurance.

1.3 Introduction to copulas.

1.3.1 Describe how a copula can be characterised 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 be investigated separately from the dependency between them.

1.3.2 Explain the 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 Describe 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 loss and their relationships.

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

2 Time series (20%)

2.1 Concepts underlying time series models.

2.1.1 Explain the concept and general properties of stationary, $I(0)$, and integrated, $I(1)$, univariate time series.

2.1.2 Explain the concept of a stationary random series.

2.1.3 Explain the concept of a filter applied to a stationary random series.

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

2.1.5 Explain the 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 Explain the concept and properties of discrete random walks and random walks with normally distributed increments, both with and without drift.

2.1.7 Explain the basic concept of a multivariate autoregressive model.

2.1.8 Explain the concept of cointegrated time series.

2.1.9 Show that certain univariate time series models have the Markov property and describe how to rearrange a univariate time series model as a multivariate Markov model.

2.2 Applications of time series models.

2.2.1 Outline 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 Describe briefly other non-stationary, non-linear time series models.

2.2.3 Describe 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%)

3.1 Describe and classify stochastic processes.

3.1.1 Define in general terms a stochastic process and in particular a counting process.

3.1.2 Classify 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.

Give examples of each type of process.

3.1.3 Describe possible applications of mixed processes.

3.1.4 Explain what is meant by the Markov property in the context of a stochastic process and in terms of filtrations.

3.2 Define and apply a Markov chain.

3.2.1 State the essential features of a Markov chain model.

3.2.2 State the Chapman–Kolmogorov equations that represent a Markov chain.

3.2.3 Calculate the stationary distribution for a Markov chain in simple cases.

3.2.4 Describe a system of frequency-based experience rating in terms of a Markov chain and describe other simple applications.

3.2.5 Describe a time-inhomogeneous Markov chain model and describe simple applications.

3.2.6 Demonstrate how 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 State the essential features of a Markov process model.

3.3.2 Define a Poisson process, derive the distribution of the number of events in a given time interval, derive the distribution of inter-event times and apply these results.

3.3.3 Derive the Kolmogorov equations for a Markov process with time-independent and time/age-dependent transition intensities.

3.3.4 Solve the Kolmogorov equations in simple cases.

3.3.5 Describe simple survival models, sickness models and marriage models in terms of Markov processes and describe other simple applications.

3.3.6 State the 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 Describe sickness and marriage models in terms of duration-dependent Markov processes and describe other simple applications.

3.3.8 Demonstrate how Markov jump processes can be used as a tool for modelling and how they can be simulated.

4 Survival models (25%)

4.1 Explain concept of survival models.

4.1.1 Describe the model of lifetime or failure time from age x as a random variable.

4.1.2 State the consistency condition between the random variable representing lifetimes from different ages.

4.1.3 Define the distribution and density functions of the random future lifetime, the survival function, the force of mortality or hazard rate, and derive relationships between them.

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

4.1.5 State the Gompertz and Makeham laws of mortality.

4.1.6 Define the curtate future lifetime from age x and state its probability function.

4.1.7 Define the symbols e_x and $\mathring{e}_x$ and derive an approximate relation between them. Define the expected value and variance of the complete and curtate future lifetimes and derive expressions for them.

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

4.2 Describe estimation procedures for lifetime distributions.

4.2.1 Describe the various ways in which lifetime data may be censored.

4.2.2 Describe the estimation of the empirical survival function in the absence of censoring and what problems are introduced by censoring.

4.2.3 Describe the 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 Describe the Nelson–Aalen estimator of the cumulative hazard rate in the presence of censoring, compute it from typical data and estimate its variance.

4.2.5 Describe models for proportional hazards and how these models can be used to estimate the impact of covariates on the hazard.

4.2.6 Describe the 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 Describe an observational plan in respect of a finite number of individuals observed during a finite period of time, and define the resulting statistics, including the waiting times.

4.3.2 Derive the likelihood function for constant transition intensities in a Markov model of transfers between states given the statistics in 4.3.1.

4.3.3 Derive maximum likelihood estimators for the transition intensities in 4.3.2 and state their asymptotic joint distribution.

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

4.4 Estimate transition intensities dependent on age (exact or census).

4.4.1 Explain the importance of dividing the data into homogeneous classes, including subdivision by age and sex.

4.4.2 Describe the principle of correspondence and explain its fundamental importance in the estimation procedure.

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

4.4.4 Calculate a central exposed to risk given the data in 4.4.3.

4.4.5 Explain how to obtain estimates of transition probabilities.

4.4.6 Explain the assumptions underlying the census approximation of waiting times.

4.4.7 Explain the concept of the rate interval.

4.4.8 Develop census formulae given age at birthday where the age may be classified as next, last or nearest relative to the birthday as appropriate, and the deaths and census data may use different definitions of age.

4.4.9 Specify the age to which estimates of transition intensities or probabilities in 4.4.8 apply.

4.5 Graduation and graduation tests.

4.5.1 Describe and apply statistical tests of the comparison 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.

4.5.2 Describe the reasons for graduating crude estimates of transition intensities or probabilities, and state the desirable properties of a set of graduated estimates.

4.5.3 Describe a test for smoothness of a set of graduated estimates.

4.5.4 Describe the process of graduation by the following methods, and state the advantages and disadvantages of each (the candidate will not be required to carry out a graduation):

- Parametric formula
- Standard table
- Spline functions.

4.5.5 Describe how the tests in 4.5.1 should be amended to compare crude and graduated sets of estimates.

4.5.6 Describe how the tests in 4.5.1 should be amended to allow for the presence of duplicate policies.

4.5.7 Carry out a comparison of a set of crude estimates and a standard table or of a set of crude estimates and a set of graduated estimates.

4.6 Mortality projection.

4.6.1 Describe the approaches to the forecasting of future mortality rates based on extrapolation, explanation and expectation, as well as their advantages and disadvantages.

4.6.2 Describe the Lee–Carter, age-period-cohort and p-spline regression models for forecasting mortality.

4.6.3 Use an appropriate computer package to apply the models in 4.6.2 to a suitable mortality data set.

4.6.4 List the main sources of error in mortality forecasts.

5 Machine learning (10%)

5.1 Explain and apply elementary principles of machine learning.

5.1.1 Explain the bias/variance trade-off and its relationship with model complexity.

5.1.2 Use cross-validation to evaluate models on unseen data, and to estimate hyper-parameters.

5.1.3 Explain how regularisation can be used to reduce overfitting in highly parameterised models.

5.1.4 Use software to apply supervised learning techniques, to solve regression and classification problems.

5.1.5 Use 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 Apply unsupervised learning techniques (principal component analysis, K-means clustering) to reduce data dimensionality, identify latent substructure and detect anomalies.

Assessment

The assessment of this subject will consist of two examinations, CS2 Paper A (CS2A) and CS2 Paper B (CS2B).

CS2A will include a number of questions with varying marks. Expect questions to be set in line with the above syllabus topic weightings and skill levels. This examination will be three hours and twenty minutes, timed and online

CS2B is a problem-based examination, focusing on computer-based data analysis and statistical modelling skills. Candidates are expected to include the R code that they have used to obtain the answers, together with the main R output produced, such as charts or tables. This examination will be one hour and fifty minutes, timed and online.

The candidate must sit both CS2A and CS2B in the same sitting, and to pass, the weighted average of the candidate's final marks achieved in both papers must exceed the pass mark for the subject. Further information on marking is available in the Studying section of the IFoA website.

Please read A Guide to CS1 and CS2 Examinations, as well as the latest version of the IFoA Examinations Handbook and IFoA Examination Regulations on the IFoA website, before sitting any IFoA examination.

END