Contents

Introduction to Actuarial Statistics (CS1 and CS2) ................................................................. 3
What is covered in CS1 and CS2? ............................................................................................. 4
How have CS1 and CS2 been developed? ............................................................................... 4
The timing of the CS1 and CS2 examinations ........................................................................ 5
What is the format of the CS examinations ............................................................................ 6
What examinations should I have passed before sitting CS1 and CS2? ............................... 6
I want to sit the examination. What do I need to do? .............................................................. 7
Professional conduct during the CS Examinations ................................................................. 7
What should I know before sitting CS1 and CS2? ................................................................ 9
R and RStudio .......................................................................................................................... 9
Word Processing ..................................................................................................................... 9
Loading data for CS1 .............................................................................................................. Error! Bookmark not defined.
Loading data for CS2 .............................................................................................................. Error! Bookmark not defined.
On completion of the examination ......................................................................................... Error! Bookmark not defined.
Final grading ............................................................................................................................ 10
Appendix 1 - CS1 Syllabus ...................................................................................................... 11
Appendix 2 - CS2 Syllabus ...................................................................................................... 22
Introduction to Actuarial Statistics (CS1 and CS2)

This guide should help answer your questions about CS1 and CS2, including:

- the format of the online examinations;
- guidance on how to take the examinations;
- administrative information; and
- technical information.

If you have any further questions that are not covered in this guide, please contact the Education Services Team: education.services@actuaries.org.uk

Important note:

Specimen CS1 and CS2 examination papers are available on the IFoA website.

https://www.actuaries.org.uk/studying/curriculum-2019
What is covered in CS1 and CS2?

Subjects CS1 and CS2 are a fundamental part of the Core Principles subjects.

CS1 provides a grounding in mathematical and statistical methods that are of relevance to actuarial work. It will equip you with the knowledge of:

- statistical distributions,
- methods to summarise data,
- the principles of statistical inference, regression models (including generalized linear models) and
- the fundamental concepts of Bayesian statistics.

CS2 builds upon CS1. It develops knowledge of and the ability to apply:

- statistical methods for risk modelling,
- time series analysis methods,
- stochastic processes (especially Markov chains and Markov jump processes),
- survival analysis (including regression methods applied to duration data) and
- graduation methods.

It also includes a high level introduction to machine learning.

Subjects CS1 and CS2 include both theory and application of the ideas using the R statistical package.

How have CS1 and CS2 been developed?

CS1 and CS2 have been influenced by external direction and guidance: the Actuarial Profession Standards (APSs), the International Standard of Actuarial Practice, the Actuaries’ Code, the regulators, Sarbanes Oxley and Solvency II requirements. A summary of these are set out for reference below. While these standards were taken into consideration when developing the syllabus for the Core Principles subjects, knowledge of these standards will not be directly tested in the CS1 and CS2 examinations.

- **Actuarial Profession Standards (APSs)**
  
The latest standards can be found on the IFoA website:  

- **The Actuaries’ Code**
  
The Actuaries Code can be found on the IFoA website:  

  The Code requires us to “perform professional duties competently and with care” and to “ensure that their communication, whether written or oral, is clear…”.

- **International Standards of Actuarial Practice (ISAP)**
  
The latest ISAPs can be found on the International Actuarial Association’s website:  
  [https://www.actuaries.org/iaa/IAA/Publications/ISAPs/IAA/Publications/05ISAPs.aspx?hkey=334b21a7-a3ac-4e0e-8294-3cbc755ab14a](https://www.actuaries.org/iaa/IAA/Publications/ISAPs/IAA/Publications/05ISAPs.aspx?hkey=334b21a7-a3ac-4e0e-8294-3cbc755ab14a)
The regulators
In reviewing a firm’s practices, the regulators expect to see acceptable standards of documentation, agreed by the firm, and documented.

Sarbanes Oxley (SOX or SarBox)
This is American legislation which governs the need for full documentation of internal processes and controls.

Solvency II
This framework for insurance companies includes emphasis on documentation and evidencing of the calculation work undertaken.

The timing of the CS1 and CS2 examinations
The dates can be found at https://www.actuaries.org.uk/studying/examination-bookings/examination-dates

The “A” papers, taken in an exam centre, are sat in local time whilst the “B” online assessments are run to UK times only, so that we can provide assistance and technical support if required.

The two elements of each assessment take place over two consecutive days. You will see these referred to as the “A” paper (sat in an examination centre) and the “B” paper (this is the online paper completed at home or in the office). When booking these two part examinations, your booking confirmation will show the date of the “A” examination. However your examination permit when issues, will show times and dates for both parts of the assessment.

Note: With effect from the April 2019 examinations the reading time that was available will now be included in the overall examination time. It is therefore up to candidates to manage their time between reading through the paper and starting to answer the questions. There is guidance on this at Hints and Tips at: https://www.actuaries.org.uk/studying/prepare-your-examinations/revision-programme
What is the format of the CS examinations?

Details of the syllabus for CS1 and CS2 are in Appendix 1 and Appendix 2. Questions for the CS1 examination and the CS2 examination will be drawn from across the syllabus. Sample examination papers can be found at:


Each of the CS1 and CS2 examination consists of two elements, Paper “A” and Paper “B”:

- Paper “A”: a 3hrs and 15mins written examination paper. This is taken in an examination centre
- Paper “B”: a 1hr and 45mins paper. This is a problem-based assessment, using R with some questions requiring candidates to analyse data loaded from external files. This can be taken at your workplace or at home, but during prescribed timings

For the problem-based assessments all candidates will need to be confident in the use of the statistical package R. There is a handy guide on Getting started with R at:


For the 2019 examinations (April and September) you will need to have access to the following. It is suggested you ensure these are available in either R Console or RStudio prior to the examinations:

<table>
<thead>
<tr>
<th>Examination</th>
<th>Version of R required</th>
<th>Version of RStudio required</th>
<th>Packages required</th>
<th>Any other requirements</th>
</tr>
</thead>
<tbody>
<tr>
<td>CS1B</td>
<td>3.5.2</td>
<td>1.1.463</td>
<td>base, stats, util, graphics</td>
<td>None</td>
</tr>
<tr>
<td>CS2B</td>
<td>3.5.2</td>
<td>1.1.463</td>
<td>fpp, markovchain, forecast, readjr</td>
<td>None</td>
</tr>
</tbody>
</table>

For detailed instructions on how to install R Console or RStudio, and how to access packages detailed above, please refer to the Getting started with R guide referenced above. In addition to the packages required for the 2019 examinations, the following packages may be helpful:

- survival
- ggplot2

In some of the examination questions you will be provided with data. These data files will have one of the following extensions: “.txt”, “.csv” or “.Rdata”. They will be provided prior to the examination. The data set(s) will also need to be loaded into R prior to the examination. For instruction on how to load data to R Console or RStudio please refer to the Getting started with R Guide referenced above.

R is continually being updated, so versions will change; sometimes this will happen quite shortly before the examination. Whilst the version of R stated above is required for sitting CS1 and CS2, version 3.5.3 may be offered at the point of download. Students can use 3.5.3 and those using 3.5.2 do not need to update the version. The IFoA will advise students before each exam session if the version of R required for CS1 or CS2 changes from that specified above.

What examinations should I have passed before sitting CS1 and CS2?

There are no pre-requisites for CS1 or CS2. However the syllabus for CS1 lists a number of foundation topics such as summarizing data sets, basic probability theory and random variables. These topics are assumed knowledge for CS1 and you may wish to ensure you are familiar with them before studying for CS1. Furthermore because CS2 builds on many of the principles covered in CS1, you may wish to ensure you are familiar with the topics covered in CS1 before studying CS2.
Candidates are advised that Formulae and Tables books may be required for these assessments. Whilst copies of these will be made available in examination centres for the A papers Candidates may use their personal copies for the online B examinations. You can purchase your copy through the e-shop at:

https://www.actuaries.org.uk/shop

I want to sit the examination. What do I need to do?

Information can be found on the IFoA website:

https://www.actuaries.org.uk/studying/examination-bookings

Professional conduct during the CS examinations

It is important to note that professional conduct is required at all times and the following will apply:

- The Assessment Regulations apply to you when taking this examination, and these can be found as Related Documents at:
  https://www.actuaries.org.uk/studying/examination-bookings/making-your-examination-booking

- Inappropriate behaviour during the CS examinations may lead to expulsion from the examination, with disciplinary consequences.
- You must not discuss the examination assignment, or disclose its contents to anyone. Failure to comply with this can result in disciplinary action being taken.
- You are not permitted to use electronic material that has been pre-prepared or copied in your examination submission (for example, pre-prepared Word or Excel documents).
- Only your first examination submissions will be accepted. Any examination submissions that are found to be modified or submitted after the examination deadline will not be accepted.

For the online examinations your attention is drawn to:

- Online examinations will be supervised via the online platform and, therefore, do not require an invigilator present at the Candidate’s chosen examination location. However where a number of candidates are taking the examination in a company office, they will need an invigilator to be provided by the company.
- Candidates must download the examination assignment and any associated materials at the start of the examination time stated. All given times are GMT/BST. Please note that the examination assignment and any associated materials will not be available at any other time.
- Candidates may refer to any written or electronic reference material provided as part of the examination day guidance documentation. All required data will normally be supplied electronically at the start of the examination time. In some circumstances, data sets may be issued ahead of the examination.
- Candidates must not use an imported e-template unless provided as part of the examination material.
- File names must include the candidate’s ARN and each Word file should also contain the candidate’s ARN as a header or footer.
- Saving work is the candidate’s responsibility. Failure to do so will not be considered a mitigating circumstance.
- It is the candidate’s responsibility to ensure that all required files are submitted within the stated timelines. Late submissions and documents found to be modified outside the examination time will be investigated and may not be marked.
- Examination submissions will only be accepted in the specified file format e.g. If MS Word is specified then .pdf submissions will not be accepted.
- Candidates are confirming by submitting the required files that all the material is entirely their own.
work and they wish this to be taken into account for the relevant assessment. Use of the internet is not permitted during the examination.
What should I know before sitting CS1 and CS2?

**R and RStudio**

- **Application**: Install R and RStudio
- **Packages**: Download and install required packages
- **R Console / RStudio**: Copy and paste appropriate output into word package
- **Formulae**: Use cell references and standard mathematical and statistical functions to generate formulae
- **Charts and graphs**: Create and label charts/graphs
- **Export graphs and charts to Word**: Export graphs and charts to Word
- **Producing output**: Check results
- **Set up output including orientation/margins/headers/footers**: Set up output including orientation/margins/headers/footers
- **Preview printing**: Preview printing
- **Print cell range**: Print cell range

**Word Processing**

- **Application**: Create/open/save/close/switch between documents
- **Adjust settings to zoom/change view**: Adjust settings to zoom/change view
- **Format document to enhance clarity including applying fonts**: Format document to enhance clarity including applying fonts
- **Text**: Insert/select/edit/copy/move/delete text
- **Use search and replace function**: Use search and replace function
- **Format text to enhance clarity including using bulleted lists**: Format text to enhance clarity including using bulleted lists
- **Charts and tables**: Insert charts and graphs from spreadsheet package, resizing if necessary
- **Create tables**: Create tables
- **Producing output**: Be aware of errors caused by Word auto-correct function
- **Set up output including orientation/margins/headers/footers**: Set up output including orientation/margins/headers/footers
- **Preview printing**: Preview printing
- **Print page range**: Print page range
On completion of the examination

At the end of the examination and within the allocated timelines, you should upload a Microsoft Word file containing your answers which should be cut and pasted from R into the Word file. You should include, where applicable, R code showing how you arrived at your answers and any relevant output from R such as charts or tables.

- You are expected to include the R code that you have used to obtain the answers, together with the main R output produced.
- When a question requires a particular numerical answer or conclusion, this should be explicitly and clearly stated, separately from, and in addition to the R output that may contain the relevant numerical information.
- Annotated plots and relevant comments should be provided when instructed to do so in the question.

Final grading

The CS examinations will assess a candidate’s overall performance over both papers. A candidate, therefore, does not have to produce a pass standard for each separate paper, but does need to demonstrate a pass standard overall. Marks for the two elements of assessment for CS1 and CS2 will be combined and a single mark awarded for each subject.
Subject CS1
Actuarial Statistics 1
Core Principles

Syllabus

for the 2019 exams

1 June 2018
CS1 – Actuarial Statistics 1

Aim

The aim of the Actuarial Statistics 1 subject is to provide a grounding in mathematical and statistical techniques that are of particular relevance to actuarial work.

Competences

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

1. describe the essential features of statistical distributions.
2. summarise data using appropriate statistical analysis, descriptive statistics and graphical presentation.
3. describe and apply the principles of statistical inference.
4. describe, apply and interpret the results of the linear regression model and generalised linear models.
5. explain the fundamental concepts of Bayesian statistics and use them to compute Bayesian estimators.

Links to other subjects

CS2 – Actuarial Statistics 2 builds directly on the material in this subject.

CM1 – Actuarial Mathematics 1 and CM2 – Actuarial Mathematics 2 apply the material in this subject to actuarial and financial modelling.

This subject assumes that a student will be competent in the following elements of Foundational Mathematics and basic statistics:

1. Summarise the main features of a data set (exploratory data analysis)

   1.1 Summarise a set of data using a table or frequency distribution, and display it graphically using a line plot, a box plot, a bar chart, histogram, stem and leaf plot, or other appropriate elementary device.
1.2 Describe the level/location of a set of data using the mean, median, mode, as appropriate.

1.3 Describe the spread/variability of a set of data using the standard deviation, range, interquartile range, as appropriate.

1.4 Explain what is meant by symmetry and skewness for the distribution of a set of data.

2 Probability

2.1 Set functions and sample spaces for an experiment and an event.

2.2 Probability as a set function on a collection of events and its basic properties.

2.3 Calculate probabilities of events in simple situations.

2.4 Derive and use the addition rule for the probability of the union of two events.

2.5 Define and calculate the conditional probability of one event given the occurrence of another event.

2.6 Derive and use Bayes’ Theorem for events.

2.7 Define independence for two events, and calculate probabilities in situations involving independence.

3 Random variables

3.1 Explain what is meant by a discrete random variable, define the distribution function and the probability function of such a variable, and use these functions to calculate probabilities.

3.2 Explain what is meant by a continuous random variable, define the distribution function and the probability density function of such a variable, and use these functions to calculate probabilities.

3.3 Define the expected value of a function of a random variable, the mean, the variance, the standard deviation, the coefficient of skewness and the moments of a random variable, and calculate such quantities.

3.4 Evaluate probabilities associated with distributions (by calculation or by referring to tables as appropriate).

3.5 Derive the distribution of a function of a random variable from the distribution of the random variable.
Syllabus topics

1. Random variables and distributions (20%)
2. Data analysis (15%)
3. Statistical inference (20%)
4. Regression theory and applications (30%)
5. Bayesian statistics (15%)

The 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 which 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 which 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 written 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 (20%)

1.1 Define basic univariate distributions and use them to calculate probabilities, quantiles and moments.

1.1.1 Define and explain the key characteristics of the discrete distributions: geometric, binomial, negative binomial, hypergeometric, Poisson and uniform on a finite set.

1.1.2 Define and explain the key characteristics of the continuous distributions: normal, lognormal, exponential, gamma, chi-square, $t$, $F$, beta and uniform on an interval.

1.1.3 Evaluate probabilities and quantiles associated with distributions (by calculation or using statistical software as appropriate).

1.1.4 Define and explain the key characteristics of the Poisson process and explain the connection between the Poisson process and the Poisson distribution.

1.1.5 Generate basic discrete and continuous random variables using the inverse transform method.

1.1.6 Generate discrete and continuous random variables using statistical software.

1.2 Independence, joint and conditional distributions, linear combinations of random variables

1.2.1 Explain what is meant by jointly distributed random variables, marginal distributions and conditional distributions.

1.2.2 Define the probability function/density function of a marginal distribution and of a conditional distribution.

1.2.3 Specify the conditions under which random variables are independent.

1.2.4 Define the expected value of a function of two jointly distributed random variables, the covariance and correlation coefficient between two variables, and calculate such quantities.

1.2.5 Define the probability function/density function of the sum of two independent random variables as the convolution of two functions.

1.2.6 Derive the mean and variance of linear combinations of random variables.
1.2.7 Use generating functions to establish the distribution of linear combinations of independent random variables.

1.3 Expectations, conditional expectations

1.3.1 Define the conditional expectation of one random variable given the value of another random variable, and calculate such a quantity.

1.3.2 Show how the mean and variance of a random variable can be obtained from expected values of conditional expected values, and apply this.

1.4 Generating functions

1.4.1 Define and determine the moment generating function of random variables.

1.4.2 Define and determine the cumulant generating function of random variables.

1.4.3 Use generating functions to determine the moments and cumulants of random variables, by expansion as a series or by differentiation, as appropriate.

1.4.4 Identify the applications for which a moment generating function, a cumulant generating function and cumulants are used, and the reasons why they are used.

1.5 Central Limit Theorem – statement and application

1.5.1 State the Central Limit Theorem for a sequence of independent, identically distributed random variables.

1.5.2 Generate simulated samples from a given distribution and compare the sampling distribution with the Normal.

2 Data analysis (15%)

2.1 Exploratory data analysis

2.1.1 Describe the purpose of exploratory data analysis.

2.1.2 Use appropriate tools to calculate suitable summary statistics and undertake exploratory data visualizations.

2.1.3 Define and calculate Pearson’s, Spearman’s and Kendall’s measures of correlation for bivariate data, explain their interpretation and perform statistical inference as appropriate.

2.1.4 Use Principal Components Analysis to reduce the dimensionality of a complex data set.
2.2 Random sampling and sampling distributions

2.2.1 Explain what is meant by a sample, a population and statistical inference.

2.2.2 Define a random sample from a distribution of a random variable.

2.2.3 Explain what is meant by a statistic and its sampling distribution.

2.2.4 Determine the mean and variance of a sample mean and the mean of a sample variance in terms of the population mean, variance and sample size.

2.2.5 State and use the basic sampling distributions for the sample mean and the sample variance for random samples from a normal distribution.

2.2.6 State and use the distribution of the $t$-statistic for random samples from a normal distribution.

2.2.7 State and use the $F$ distribution for the ratio of two sample variances from independent samples taken from normal distributions.

3 Statistical inference (20%)

3.1 Estimation and estimators

3.1.1 Describe and apply the method of moments for constructing estimators of population parameters.

3.1.2 Describe and apply the method of maximum likelihood for constructing estimators of population parameters.

3.1.3 Define the terms: efficiency, bias, consistency and mean squared error.

3.1.4 Define and apply the property of unbiasedness of an estimator.

3.1.5 Define the mean square error of an estimator, and use it to compare estimators.

3.1.6 Describe and apply the asymptotic distribution of maximum likelihood estimators.

3.1.7 Use the bootstrap method to estimate properties of an estimator.
3.2 Confidence intervals

3.2.1 Define in general terms a confidence interval for an unknown parameter of a distribution based on a random sample.

3.2.2 Derive a confidence interval for an unknown parameter using a given sampling distribution.

3.2.3 Calculate confidence intervals for the mean and the variance of a normal distribution.

3.2.4 Calculate confidence intervals for a binomial probability and a Poisson mean, including the use of the normal approximation in both cases.

3.2.5 Calculate confidence intervals for two-sample situations involving the normal distribution, and the binomial and Poisson distributions using the normal approximation.

3.2.6 Calculate confidence intervals for a difference between two means from paired data.

3.2.7 Use the bootstrap method to obtain confidence intervals.

3.3 Hypothesis testing and goodness of fit

3.3.1 Explain what is meant by the terms null and alternative hypotheses, simple and composite hypotheses, type I and type II errors, test statistic, likelihood ratio, critical region, level of significance, probability-value and power of a test.

3.3.2 Apply basic tests for the one-sample and two-sample situations involving the normal, binomial and Poisson distributions, and apply basic tests for paired data.

3.3.3 Apply the permutation approach to non-parametric hypothesis tests.

3.3.4 Use a chi-square test to test the hypothesis that a random sample is from a particular distribution, including cases where parameters are unknown.

3.3.5 Explain what is meant by a contingency (or two-way) table, and use a chi-square test to test the independence of two classification criteria.

4 Regression theory and applications (30%)

4.1 Linear regression

4.1.1 Explain what is meant by response and explanatory variables.

4.1.2 State the simple regression model (with a single explanatory variable).
4.1.3 Derive the least squares estimates of the slope and intercept parameters in a simple linear regression model.

4.1.4 Use appropriate software to fit a simple linear regression model to a data set and interpret the output.
- Perform statistical inference on the slope parameter.
- Describe the use of measures of goodness of fit of a linear regression model.
- Use a fitted linear relationship to predict a mean response or an individual response with confidence limits.
- Use residuals to check the suitability and validity of a linear regression model.

4.1.5 State the multiple linear regression model (with several explanatory variables).

4.1.6 Use appropriate software to fit a multiple linear regression model to a data set and interpret the output.

4.1.7 Use measures of model fit to select an appropriate set of explanatory variables.

4.2 Generalised linear models

4.2.1 Define an exponential family of distributions. Show that the following distributions may be written in this form: binomial, Poisson, exponential, gamma, normal.

4.2.2 State the mean and variance for an exponential family, and define the variance function and the scale parameter. Derive these quantities for the distributions above.

4.2.3 Explain what is meant by the link function and the canonical link function, referring to the distributions above.

4.2.4 Explain what is meant by a variable, a factor taking categorical values and an interaction term. Define the linear predictor, illustrating its form for simple models, including polynomial models and models involving factors.

4.2.5 Define the deviance and scaled deviance and state how the parameters of a generalised linear model may be estimated. Describe how a suitable model may be chosen by using an analysis of deviance and by examining the significance of the parameters.

4.2.6 Define the Pearson and deviance residuals and describe how they may be used.
4.2.7 Apply statistical tests to determine the acceptability of a fitted model: Pearson’s chi-square test and the likelihood ratio test

4.2.8 Fit a generalised linear model to a data set and interpret the output.

5 Bayesian statistics (15%)

5.1 Explain the fundamental concepts of Bayesian statistics and use these concepts to calculate Bayesian estimators.

5.1.1 Use Bayes’ theorem to calculate simple conditional probabilities.

5.1.2 Explain what is meant by a prior distribution, a posterior distribution and a conjugate prior distribution.

5.1.3 Derive the posterior distribution for a parameter in simple cases.

5.1.4 Explain what is meant by a loss function.

5.1.5 Use simple loss functions to derive Bayesian estimates of parameters.

5.1.6 Explain what is meant by the credibility premium formula and describe the role played by the credibility factor.

5.1.7 Explain the Bayesian approach to credibility theory and use it to derive credibility premiums in simple cases.

5.1.8 Explain the empirical Bayes approach to credibility theory and use it to derive credibility premiums in simple cases.

5.1.9 Explain the differences between the two approaches and state the assumptions underlying each of them.

Assessment

Assessment consists of a combination of a computer-based data analysis and statistical modelling assignment and a three-hour written examination.

END
Subject CS2A
Risk Modelling and Survival Analysis
Core Principles

Syllabus

for the 2019 exams

1 June 2018
CS2 — Risk Modelling and Survival Analysis

Core Principles

Aim

The aim of the Actuarial Statistics 2 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 student 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 student is competent with the material covered in CS1 – Actuarial Statistics 1 and the required knowledge for that subject.

CM1 – Actuarial Mathematics 1 and CM2 – Actuarial Mathematics 2 apply the material in this subject to actuarial and financial modelling.

Topics in this subject are further built upon in SP1 – Health and Care Principles, SP7 – General Insurance Reserving and Capital Modelling Principles, SP8 – General Insurance Pricing Principles and SP9 – Enterprise Risk Management Principles.
CS2

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%)

The 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 which 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 which 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 written 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 which 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 dataset 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 which 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 might 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

and 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 \( p_x \) and \( q_x \) 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 \( \overline{e}_x \) 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 might 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, including in the single decrement model the actuarial estimate based on the simple adjustment to the central exposed to risk.

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:

• parametric formula
• standard table
• spline functions

(The student will not be required to carry out a graduation.)

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, and 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 dataset.

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 main branches of machine learning and describe examples of the types of problems typically addressed by machine learning.

5.1.2 Explain and apply high-level concepts relevant to learning from data.
5.1.3 Describe and give examples of key supervised and unsupervised machine learning techniques, explaining the difference between regression and classification and between generative and discriminative models.

5.1.4 Explain in detail and use appropriate software to apply machine learning techniques (e.g. penalised regression and decision trees) to simple problems.

5.1.5 Demonstrate an understanding of the perspectives of statisticians, data scientists, and other quantitative researchers from non-actuarial backgrounds.

Assessment

Combination of a computer based data analysis and statistical modelling assignment and a three hour written examination.

END