This manual includes GOAL
Customizable, versatile online exam question bank. Thousands of questions!

Access your exclusive ***StudyPlus+*** bonus content:
GOAL | Flashcards | Career Guides

a/s/m
Exam MAS-II Study Manual

Abraham Weishaus, Ph.D., F.S.A., C.F.A., M.A.A.A.

a/s/m
Actuarial Study Materials
Learning Made Easier

Introduction

Welcome to Exam MAS-II!

This exam covers four subjects: credibility, linear mixed models, Bayesian methods, and statistical learning. The syllabus lists them in that order, and the exam questions past the second question are in that order. (The first two questions on each exam are case study questions, and relate to linear mixed models.) While you can study the subjects in any order (with some care), that order is the most logical order.

The first three topics are related in that they all deal with how to predict given two drivers. One of the drivers is always the current data specifically related to your group. The other driver may be judgment, previous knowledge, or data from the entire universe. In credibility, we are balancing data from the group we want to rate with general data from the entire universe, or with our previous rate. In linear mixed models, we are balancing data from each cluster with data from all clusters in order to predict for each cluster. In Bayesian methods, we are balancing data with our general knowledge, which may include thoughts like "the result must be positive" or "the results are probably in a specific range", and we can even specify how much weight should be given to our general knowledge.

Credibility calculations have been done for a long time; the computations can be done on a calculator. But linear mixed models and Bayesian methods have only become practical relatively recently with the availability of fast computers.

Statistical learning is a totally separate subject, except that "cross-entropy" is referred to both there and in Bayesian methods.

What are the weights of these subjects on the exam? Table 1 answers that question.

Table 1: Weights of the Four Subjects on MAS-II

Subject	Syllabus Weight	Question Counts		
		F18	S19	F19
Credibility	5–15%	4	6	4
Linear mixed models	10–30%	11	10	11
Bayesian methods	45–65%	20	21	22
Statistical learning	10–20%	7	5	5
Total		42	42	42

Note that through the end of 2020, the weight on linear mixed models was 5% higher (15–35%) and the weight on Bayesian methods was 5% lower (40–60%). The question count followed the old syllabus weights. But future exams will have about 2 more questions on Bayesian methods and 2 fewer questions on linear mixed models.

Each of the old exams had 42 questions. The first two questions were case study questions. For these questions, you were given supplementary material, a packet of 20–40 pages, that discussed the situation that was studied, followed by preliminary analysis mostly through plots, followed by outputs from many runs of alternative models using R. The outputs include tables and various diagnostic plots.

About the exam

At this writing, the syllabus for the exam is at

https://www.casact.org/sites/default/files/2021-11/ExamMASII_Syllabus.pdf.

Download the tables from

https://www.casact.org/sites/default/files/2021-03/masii_tables.pdf.

These tables include distribution tables that you also got for Exam MAS-I and various statistical tables such as normal distribution and chi-square distribution. You will get these tables at the exam. But on recent exams they are hardly needed. For example, on the Fall 2019 exam, if you knew the Poisson and binomial distributions and the mean of a gamma distribution, you would not need the tables.

CAS exams frequently have defective questions. For example, if you check the answer key for the Fall 2019 exam, you will see multiple answers were allowed for two questions. In addition, one question's assumptions were impossible; it stated that a mode for a gamma distribution was greater than its mean, which is impossible. Just be ready for such questions and do your best. Try to figure out what the examiner really meant. And realize that they'll probably give credit for any reasonable interpretation. With the move to computer-based testing (CBT), questions will probably be proofread more closely, and defects may show up and be fixed while questions are still pilot questions. So I expect defective questions to be less common in the future.

Talking about CBT, you get a one-sheet spreadsheet with Excel functions at the exam. See

https://wsr.pearsonvue.com/testtaker/common/StartTestLaunch.htm?clientCode=CAS&seriesCode=Spreadsheet_SAMPLE&languageCode=ENU

The Excel functions available are listed here:

https://home.pearsonvue.com/Clients/Casualty-Actuarial-Society-(CAS)/Spreadsheet-Function-List.aspx

The CAS website in early 2021 said that the questions on the Spring 2021 exam do not take advantage of the additional spreadsheet capabilities but will be similar to prior paper and pencil exams. But in the future they may have more sophisticated test questions and use larger datasets. This manual assumes that for the 2022 exams the same will hold.

As a result of CBT, "the order of questions will be randomly presented within the various sections of the syllabus". I interpret this to mean that the questions for each section of the syllabus will still be kept together, but they will be in a random order within each section.

This manual

This manual has everything you need to score a 10 on your exam. But (with the exception of credibility) it does not have enough material to do the programming that would be necessary at your job. For that, you need the textbooks, plus a good knowledge of R or another statistical programming language.

Based on Table 1, credibility is officially about 10% of the ex am. In reality, it is closer to 20% of the exam, since exams typically have about 3–4 questions in their Bayesian methods section that are covered in the credibility part of this manual.

However, unlike the other subjects on this exam, credibility has been tested for many decades. As a result, there are tons of old exam questions available. I've taken the old material, discarded about 45% of it, and updated the notation to match the syllabus textbook, but it is still a lot of material relative to its weight on the exam. In the introduction to credibility, I give guidance on which lessons are most important. That way if you have a lot of time and want to make sure you get all 7–8 credibility questions right, you have everything you need to do that. But if you're aiming to get a good score on the exam and don't mind missing 1 or 2 credibility questions, you can just study the important parts, do maybe half the exercises, and move on.

While we're talking about the importance of different lessons in this manual, here is a table with the number of old exam questions from the three released exams coming from each lesson.[1]

[1] Lesson 31 is a new topic, not on the syllabus before 2021, hence omitted from the table.

Table 2: Released Exam Question Counts by Lesson

Lesson	#	Lesson	#	Lesson	#	Lesson	#
1	1	11	1	21	4	32	9
2	1	12	3	22	0	33	11
3	2	13	0	23	1	34	7
4	6	14	2	24	3	35	2
5	1	15	1	25	1	36	2
6	0	16	2	26	6	37	8
7	2	17	6	27	0	38	3
8	3	18	5	28	8	39	5
9	0	19	3	29	3	40	5
10	0	20	3	30	2	41	4

In reading this table, keep in mind:

- In this manual, the lessons associated with each subject are

Credibility	1–16
Linear mixed models	17–27
Bayesian methods	28–37
Statistical learning	38–41

- Exam questions may require material from many lessons. This table classifies old exam questions based on the highest lesson number they are based on. For example, the questions associated with Lesson 26 are the two case study questions on each exam, but solving these requires understanding a lot of the linear mixed model material from earlier lessons.

New for the second edition of this manual

This edition of the manual was updated in accordance with the updated CAS syllabus for Exam MAS-II. That syllabus updated topics on Linear Mixed Models and Bayesian Methods.

For Linear Mixed Models, the syllabus states that the only language you are responsible for is R; you need not know Stata, HLM, SAS, or SPSS. Moreover, you are not expected to code in R; you just need to be able to understand typical outputs and reports. You are not responsible for knowing the peculiarities of the different languages. The syllabus explicitly specifies the sections of the textbook you may skip. For the most part, the first edition of the manual did not cover any of them, with one exception: it discussed denominator degrees of freedom in R. That discussion has been removed. The manual still does not cover some obscure sections of the textbook that are theoretically still on the syllabus. If you get a question on Kronecker products (or any other obscure topic not covered in this manual), make sure to contact me!

A second edition of the textbook for Bayesian Methods, *Statistical Rethinking*, was released. The updated syllabus includes much of the new material in that textbook, and that is perhaps the reason that the weight on this topic was increased by 5%, to 45%–65%; the majority of the exam will be on this topic! Topics removed from the second edition are DIC and model averaging. Yet the CAS syllabus, which insists that you use the second edition, still lists these topics. About the only question they could ask on model averaging (and you have to read the endnote to answer it) is

> You are given the following statements regarding model averaging.
>
> I. Model averaging is a family of methods for combining predictions of multiple models.
>
> II. Model averaging is focused on prediction rather than inference.
>
> III. For the sake of space, *Statistical Rethinking* doesn't cover it.
>
> Determine which of the above statements is true.

(Correct answer is: all three are true.) But I doubt the exam would have a silly question like this.

Many topics were added, and the author changed his opinions and methods for many of the topics covered in the first edition. New topics are:

1. Examination of priors for reasonableness. The prior predictive is studied to see whether it presents a reasonable distribution of the response based on general knowledge. In many cases, priors that were used in the first edition are rejected as ridiculous.

2. B-splines. Fitting basis splines to distributions. You had splines on a non-Bayesian basis in your studies for Exam MAS-I.

3. A comprehensive framework for spotting confounds in multiple regressions using directed acyclic graphs (DAGs).

4. Modeling categorical variables with indexing rather than dummy variables was already mentioned in the first edition, but in the second edition use of dummy variables is completely rejected.

5. PSIS, an alternative to WAIC and to the no-longer-discussed DIC.

6. A section on interactions that was already in the first edition was added to the CAS syllabus. In exchange, the CAS dropped a section introducing generalized linear models. You already know what those are from your Exam MAS-I studies, so you can proceed immediately to the Bayesian method for GLMs with no introduction.

7. Robust regression using Student's T priors instead of Gaussian priors, to handle outliers.

8. Hamiltonian Monte Carlo is more fully explained. And an additional diagnostic plot, the trank plot, is discussed.

9. Ordered categorical predictors.

10. The section on noncentered parametrization was expanded. Noncentered parametrization is one way to handle divergent transitions.

Most of these topics are beyond hand-computation and require `R` to implement. Most exam questions on these topics are likely to be conceptual. I've provided a small number of exercises and some practice exam questions on these topics.

Exams are no longer going to be released, so I may never know what the CAS is testing on. I've done my best to cover all topics that may appear on the exam.

New for the third edition of this manual

The principal components analysis lesson was rewritten in a clearer way.

The CAS revamped their website. In the process of revamping, they removed the Fall 2019 exam case study. Previously, they had removed the Fall 2018 exam case study, and unfortunately I didn't download it before it was removed. The Spring 2019 exam includes the case study as part of the pdf, and is still there, and the sample case study is still there.

This manual includes a new case study to replace the Fall 2019 case study. But it also includes the Fall 2019 exam case study in the Appendix, along with original questions in Lesson 26. Generally the CAS has allowed us to use their old exam questions, and the case study is part of the old exam questions. But if they object, I will have to remove it, along with the associated questions.

Acknowledgements

I would like to thank the SOA and CAS for allowing me to use their old and sample exam questions.

The creators of TEX, LATEX, and its multitude of packages all deserve thanks for making possible the professional typesetting of this mathematical material.

I would like to thank Nao Mimoto for proofreading the linear mixed models and Bayesian methods parts of this manual.

I would like to thank the readers who submitted errata: Tiffany Chang, Carrie Cui, Christopher Ng Chieng Hin, Michelle Lau, Tzong Her Lee, Jonathan Lim, Trevor McElhinney, Hannah Park, Rachel Ruble, 4Casey Schnakenberg, Amy Schroeder, Yuliya Semibratova, Alex Shepard, Allen Sherpa, Yongkang Wang, Joe Wooton.

Errata

Please report all errors you find in this manual to the author. You may send them to the publisher at mail@studymanuals.com or directly to me at errata@aceyourexams.net.

An errata list will be posted at http://errata.aceyourexams.net. Check this errata list frequently.

Lesson 2

Classical Credibility: Non-Poisson Frequency

Reading: Tse 6.4

Undoubtedly you're scratching your head. What is this lesson doing here? Tse Section 6.4 is not on the syllabus! OK, feel free to skip this lesson, but be aware that the Fall 2019 exam had a question on it. (See exercise 2.10.) This question, in fact, is very similar to Tse Example 6.12, which is in Section 6.4.

As we saw in the previous lesson (formula (1.1)), the general formula for the standard for full credibility in *exposure units* is

$$e_F = \lambda_F \left(\frac{\sigma}{\mu}\right)^2$$

where e_F is the standard *measured in exposure units, not claims*, μ is the mean and σ is the standard deviation of the item you are measuring the credibility of. To obtain n_F, the standard measured in expected claims, we multiply e_F by the mean frequency of claims, μ_N. This general formula is used to determine the number of exposure units (e.g., policy years) needed for full credibility of the pure premium if you're only given the mean and variance of aggregate claims, but not the separate means and variances of frequency and severity.

If you are establishing a standard for full credibility of claim frequency in terms of the number of exposures, formula (1.1) translates into

$$e_F = \lambda_F \left(\frac{\sigma_N^2}{\mu_N^2}\right) \tag{2.1}$$

where μ_N is the mean of frequency and σ_N^2 is the variance of frequency. To express this standard in terms of number of expected claims, we multiply both sides by μ_N to obtain

$$n_F = \lambda_F \left(\frac{\sigma_N^2}{\mu_N}\right) \tag{2.2}$$

Notice how this generalizes the formula of the previous lesson for Poisson-distributed claim frequency, where $\sigma_N^2 = \mu_N$.

If you are establishing a standard for full credibility of aggregate loss or pure premium, assuming that claim counts and claim sizes are independent, the formula for the (expected) number of claims needed for full credibility can be derived as follows. The mean of pure premium is

$$\mathbf{E}[S] = \mathbf{E}[N]\,\mathbf{E}[X] = \mu_N \mu_X$$

By the compound variance formula, equation (1.5):

$$\mathrm{Var}(S) = \mathbf{E}[N]\,\mathrm{Var}(X) + \mathrm{Var}(N)\,\mathbf{E}[X]^2 = \mu_N \sigma_X^2 + \sigma_N^2 \mu_X^2$$

From formula (1.1), noting that $n_F = \mu_N e_F$:

$$\begin{aligned} n_F &= \mu_N \lambda_F \frac{\mu_N \sigma_X^2 + \sigma_N^2 \mu_X^2}{\mu_N^2 \mu_X^2} \\ &= \lambda_F \left(\frac{\sigma_N^2}{\mu_N} + C_X^2\right) \end{aligned} \tag{2.3}$$

Table 2.1: Classical credibility formulas

	Credibility for		
Experience expressed in	Number of claims	Claim size (severity)	Aggregate losses/ Pure premium
Policyholder-years	$\lambda_F\left(\dfrac{\sigma_N^2}{\mu_N^2}\right)$	N/A	$\lambda_F\left(\dfrac{\sigma_N^2}{\mu_N^2}+\dfrac{\sigma_X^2}{\mu_X^2\mu_N}\right)$
Number of claims	$\lambda_F\left(\dfrac{\sigma_N^2}{\mu_N}\right)$	$\lambda_F\left(\dfrac{\sigma_X^2}{\mu_X^2}\right)$	$\lambda_F\left(\dfrac{\sigma_N^2}{\mu_N}+\dfrac{\sigma_X^2}{\mu_X^2}\right)$

Notice the asymmetry: the denominator for frequency is μ_N not squared, whereas $C_X^2 = \sigma_X^2/\mu_X^2$. This standard can be expressed in terms of exposure units by dividing n_F by μ_N.

Example 2A You are given:

- Claim counts follow a negative binomial distribution with $r = 2$ and $\beta = 0.5$.
- Claim sizes have coefficient of variation equal to 2.5.
- Claim counts and claim sizes are independent.

The standard for full credibility of aggregate losses is set so that actual aggregate losses are within 5% of expected 95% of the time.

Determine the number of expected claims needed for full credibility.

Solution: The ratio of the variance of a negative binomial, $r\beta(1+\beta)$, to its mean, $r\beta$ is $1+\beta$, which is 1.5 here. Using formula (2.3),

$$n_F = \left(\frac{1.96}{0.05}\right)^2 (1.5 + 2.5^2) = \boxed{\mathbf{11{,}909}}$$

□

From this example, you see that σ_N^2/μ_N for a negative binomial is $1+\beta$. It may also be useful to know that for a binomial, $\sigma_N^2/\mu_N = 1-q$.

A summary of the formulas for all possible combinations of experience units used and what the credibility is for is shown in Table 2.1. To make the formulas parallel, I've avoided using the coefficient of variation. The two most common formulas are shaded.

Exercises

2.1. The methods of classical credibility are used. The standard for full credibility is that the item measured should be within $100k\%$ of the true mean with probability p.

Order the following items from lowest to highest. In each case, assume the distribution is non-degenerate (in other words, that the random variable is not a constant).

I. Number of losses for full credibility of individual losses if losses follow an exponential distribution.

II. Number of expected claims for full credibility of claim counts if claim counts follow a binomial distribution with $m = 1$.

III. Number of expected claims for full credibility of aggregate losses if claim counts follow a Poisson distribution and loss sizes follow a two-parameter Pareto distribution.

A. I < II < III B. I < III < II C. II < I < III D. II < III < I E. III < II < I

Exercises continue on the next page . . .

2.2. Claim size follows a two parameter Pareto distribution with parameters $\alpha = 3.5$ and θ. The full credibility standard for claim size is set so that actual average claim size is within 5% of expected claim size with probability 98%.

Determine the number of claims needed for full credibility.

2.3. Claim size follows an inverse Gaussian distribution with parameters $\mu = 1000$, $\theta = 500$. You set a standard for full credibility of claim size, using the methods of classical credibility, so that expected claim size is within $100k$% of actual claim size with probability 90%. Under this standard, 1500 claims are needed for full credibility.

Determine k.

2.4. You are given the following information for a risk.

Claim frequency: mean = 0.2, variance = 0.3
Claim severity: gamma distribution with $\alpha = 2$, $\theta = 10{,}000$.

Using the methods of classical credibility, determine the number of expected claims needed so that aggregate claims experienced are within 5% of expected claims with probability 90%.

A. Less than 1800
B. At least 1800, but less than 1900
C. At least 1900, but less than 2000
D. At least 2000, but less than 2100
E. At least 2100

2.5. For a certain coverage, claim frequency has a negative binomial distribution with $\beta = 0.25$. The full credibility standard is set so that the actual number of claims is within 6% of the expected number with probability 95%.

Determine the number of expected claims needed for full credibility.

2.6. **[4B-F94:15]** (3 points) You are given the following:

- Y represents the number of independent homogeneous exposures in an insurance portfolio.
- The claim frequency *rate* per exposure is a random variable with mean = 0.025 and variance = 0.0025.
- A full credibility standard is devised that requires the observed sample frequency *rate* per exposure to be within 5% of the expected population frequency rate per exposure 90% of the time.

Determine the value of Y needed to produce full credibility for the portfolio's experience.

A. Less than 900
B. At least 900, but less than 1500
C. At least 1500, but less than 3000
D. At least 3000, but less than 4500
E. At least 4500

2.7. **[4-F04:21, STAM Sample Question #148]** You are given:

- The number of claims has probability function:

$$p(x) = \binom{m}{x} q^x (1-q)^{m-x} \qquad x = 0, 1, 2, \ldots, m$$

- The actual number of claims must be within 1% of the expected number of claims with probability 0.95.
- The expected number of claims for full credibility is 34,574.

Determine q.

A. 0.05 B. 0.10 C. 0.20 D. 0.40 E. 0.80

Exercises continue on the next page . . .

2.8. **[4-F00:14]** For an insurance portfolio, you are given:

- For each individual insured, the number of claims follows a Poisson distribution.
- The mean claim count varies by insured, and the distribution of mean claim counts follows a gamma distribution.
- For a random sample of 1000 insureds, the observed claim counts are as follows:

Number Of Claims, n	0	1	2	3	4	5
Number Of Insureds, f_n	512	307	123	41	11	6

$$\sum n f_n = 750 \qquad \sum n^2 f_n = 1494$$

- Claim sizes follow a Pareto distribution with mean 1500 and variance 6,750,000.
- Claim sizes and claim counts are independent.
- The full credibility standard is to be within 5% of the expected aggregate loss 95% of the time.

Determine the minimum number of insureds needed for the aggregate loss to be fully credible.

A. Less than 8300
B. At least 8300, but less than 8400
C. At least 8400, but less than 8500
D. At least 8500, but less than 8600
E. At least 8600

2.9. **[C-S05:2, STAM Sample Question #173]** You are given:

- The number of claims follows a negative binomial distribution with parameters r and $\beta = 3$.
- Claim severity has the following distribution:

Claim Size	Probability
1	0.4
10	0.4
100	0.2

- The number of claims is independent of the severity of claims.

Determine the expected number of claims needed for aggregate losses to be within 10% of expected aggregate losses with 95% probability.

A. Less than 1200
B. At least 1200, but less than 1600
C. At least 1600, but less than 2000
D. At least 2000, but less than 2400
E. At least 2400

Exercises continue on the next page . . .

2.10. [**MAS-II-F19:4**] You are given the following parameters.

- Assume the full-credibility standard using classical credibility is based on $P = 0.05$ and $k = 0.02$.
- The expected claim frequency per exposure unit is 0.03.
- W is the full-credibility standard for claim frequency in exposure units assuming Poisson claim frequency.
- V is the full-credibility standard for claim frequency in exposure units assuming Bernoulli claim frequency.

Calculate $|W - V|$.

A. Less than 4,000
B. At least 4,000, but less than 6,000
C. At least 6,000, but less than 8,000
D. At least 8,000, but less than 10,000
E. At least 10,000

Solutions

2.1. For exponential X with mean θ, the variance is θ^2 so the coefficient of variation squared is $\theta^2/\theta^2 = 1$ and the standard for full credibility, using the general formula, is $n_0 C^2 = n_0$.

For a binomial with parameters $m = 1$ and q, the coefficient of variation squared is $q(1-q)/q^2 = (1-q)/q$. The number of exposures for full credibility, by the general formula, is $n_0(1-q)/q$, and the number of expected claims per exposure is q, so the number of expected claims needed for full credibility is $n_0(1-q)$. Since $q > 0$, this is less than I. (If $q = 0$, the distribution is degenerate; the random variable is the constant 0.)

For a two-parameter Pareto, the standard for full credibility in terms of expected claims is $2n_0(\alpha-1)/(\alpha-2)$, based on Table 1.2. This is greater than $2n_0$, hence greater than I. **(C)**

2.2. Since we want credibility for *severity* using number of claims (exposure), formula (1.7) is the appropriate one. We calculate C_X^2.

$$\mu_X = \frac{\theta}{2.5}$$
$$\sigma_X^2 = \frac{2\theta^2}{(2.5)(1.5)} - \frac{\theta^2}{2.5^2} = \frac{28}{75}\theta^2$$
$$C_X^2 = \left(\frac{\sigma}{\mu}\right)^2 = \frac{28}{75}\left(\frac{25}{4}\right) = \frac{7}{3}$$

Using the formula, we conclude

$$n_F = \left(\frac{2.326}{0.05}\right)^2\left(\frac{7}{3}\right) = \boxed{\mathbf{5049.6}}$$

2.3. Since we want credibility for *severity* using number of claims (exposure), formula (1.7) is the appropriate one. To back out the k, we must first calculate C_X.

$$\mu_X = 1000$$
$$\mu_X = \frac{10^9}{500}$$
$$C_X^2 = 2$$

Now we equate 1500 to e_F.

$$1500 = \left(\frac{1.645}{k}\right)^2(2)$$

$$1500k^2 = (1.645^2)(2)$$
$$750k^2 = 1.645^2$$
$$k = \frac{1.645}{\sqrt{750}} = \boxed{0.06}$$

2.4. We want credibility for aggregate claims using number of expected claims as the basis. With separate information on frequency and severity, formula (2.3) applies. We must calculate C_X.

$$\mu_X = 2(10{,}000) = 20{,}000$$
$$\mu_X = 2(10{,}000^2) = 2 \cdot 10^8$$
$$C_X^2 = \frac{2 \cdot 10^8}{20{,}000^2} = \frac{1}{2}$$

We now apply formula (2.3).

$$n_F = 1082.41\left(\frac{0.3}{0.2} + \frac{1}{2}\right) = \boxed{2164.82} \qquad \textbf{(E)}$$

2.5. We want credibility for frequency using number of expected claims as the basis. Formula (2.2) applies.

$$\frac{\sigma_N^2}{\mu_N} = 1.25$$
$$n_F = \left(\frac{1.96}{0.06}\right)^2 (1.25) = \boxed{1333.89}$$

2.6. We want credibility for frequency using exposures as the basis. Formula (2.2) applies.

$$n_F = 1082.41\left(\frac{0.0025}{0.025}\right) = 108.241$$
$$Y = e_F = \frac{108.241}{0.025} = \boxed{4329.64} \qquad \textbf{(D)}$$

2.7. We have

$$\lambda_F = \left(\frac{1.96}{0.01}\right)^2 \left(\frac{\sigma^2}{\mu}\right)$$

and for a binomial, $\sigma^2 = mq(1-q)$ and $\mu = mq$ so the quotient $\sigma^2/\mu = 1 - q$. Then

$$196^2(1-q) = 34{,}574$$
$$38{,}416(1-q) = 34{,}574$$
$$q = 1 - \frac{34{,}574}{38{,}416} = \boxed{0.1} \qquad \textbf{(B)}$$

2.8. The coefficient of variation for severity squared is $6{,}750{,}000/1500^2 = 3$. For frequency, we use the summary statistics to estimate the variance over the mean. Estimated mean is $750/1000 = 0.75$. Estimated variance is $1494/1000 - 0.75^2 = 0.9315$. If you wish, you can multiply this by 1000/999 (so that the sample variance is divided by $n-1$ instead of by n), but it hardly makes a difference. So we have

$$n_F = \left(\frac{1.96}{0.05}\right)^2 \left(\frac{0.9315}{0.75} + 3\right) = 6518.43$$
$$e_F = \frac{6518.43}{0.75} = \boxed{8691.24} \qquad \textbf{(E)}$$

2.9. The variance of number of claims divided by the mean is $1+\beta=4$.

The mean of claim size is $0.4(1)+0.4(10)+0.2(100)=24.4$. The second moment is $0.4(1)+0.4(100)+0.2(10{,}000)=2040.4$. The variance is then $2040.4-24.4^2=1445.04$, and the coefficient of variation squared is $1445.04/24.4^2=2.4272$. The answer is then

$$n_F=\left(\frac{1.96}{0.1}\right)^2(4+2.4272)=\boxed{2469} \qquad \textbf{(E)}$$

2.10. Note that the textbook uses P to mean "premium" and uses α to mean the probability parameter. *Do not assume CAS exam questions are error-free or use notation consistent with the syllabus textbooks!*

First calculate n_0.

$$n_0=\left(\frac{z_{0.975}}{0.02}\right)^2=\left(\frac{1.96}{0.02}\right)^2=9604$$

For Poisson frequency, this represents number of expected claims needed; divide it by 0.03 to obtain exposure units.

$$W=\frac{9604}{0.03}$$

For binomial (with $m=1$, or Bernoulli) frequency, first multiply by variance divided by mean to obtain expected claims needed, then divide by 0.03 to obtain exposure units. So multiply by $(0.03)(1-0.03)/0.03^2$.

$$V=9604\left(\frac{1-0.03}{0.03}\right)=\frac{9604}{0.03}-9604$$

Then $W-V=\boxed{9604}$. **(D)**

Lesson 3

Classical Credibility: Partial Credibility

Reading: Tse 6.3

When there is inadequate experience for full credibility, then the credibility estimate, which we will sometimes call the *credibility premium*, or P_C, is calculated by

$$P_C = Z\bar{X} + (1 - Z)M \tag{3.1}$$

where M is the manual premium, the premium initially assumed if there is no credibility, and Z is the credibility factor. For calculator purposes, it is easier to use this formula in the form

$$P_C = M + Z(\bar{X} - M) \tag{3.2}$$

since you don't need any memory, and entering M twice is likely to be easier than entering Z, which is usually a long decimal, twice. This alternative form is also intuitive; you are modifying M by adding the difference between actual experience and M, multiplied by the credibility assigned to the experience.

We want to determine Z.

We saw in the story of Ventnor Manufacturing at the beginning of Lesson 1 that to multiply the variance of the results by α, we must multiply the results by $\sqrt{\alpha}$. Therefore, the credibility factor for n expected claims is

$$Z = \sqrt{\frac{n}{n_F}} \tag{3.3}$$

where n_F is the number of expected claims needed for full credibility. The corresponding square root rule would apply to expressing credibility in exposures in terms of e_F, or credibility in terms of aggregate claims in terms of the amount needed for full credibility:

$$Z = \sqrt{\frac{e}{e_F}}$$

The partial credibility function is concave down; it grows rapidly for small numbers, then slows down. Figure 3.1 illustrates the curve if we assume 1082 claims are needed for full credibility.

Let's see how the Ventnor case fits into this formula. We established on page 7 that 1125 expected claims were needed for full credibility. We have 160 claims. Therefore $Z = \sqrt{160/1125} = 0.3771$, which matches the result we initially computed on page 3.

Quiz 3-1 If 250 expected claims result in 50% credibility, how many expected claims are needed for 20% credibility?

Example 3A (Version of 4B-S91:23) (1 point) Claim counts for a group follow a Poisson distribution. The standard for full credibility is 19,544 expected claims. We observe 6000 claims and a total loss of 15,600,000 for a group of insureds.

If our prior estimate of the total loss is 16,500,000, determine the classical credibility estimate of the total loss for the group of insureds.

A. Less than 15,780,000
B. At least 15,780,000, but less than 15,870,000
C. At least 15,870,000, but less than 15,960,000
D. At least 15,960,000, but less than 16,050,000
E. At least 16,050,000

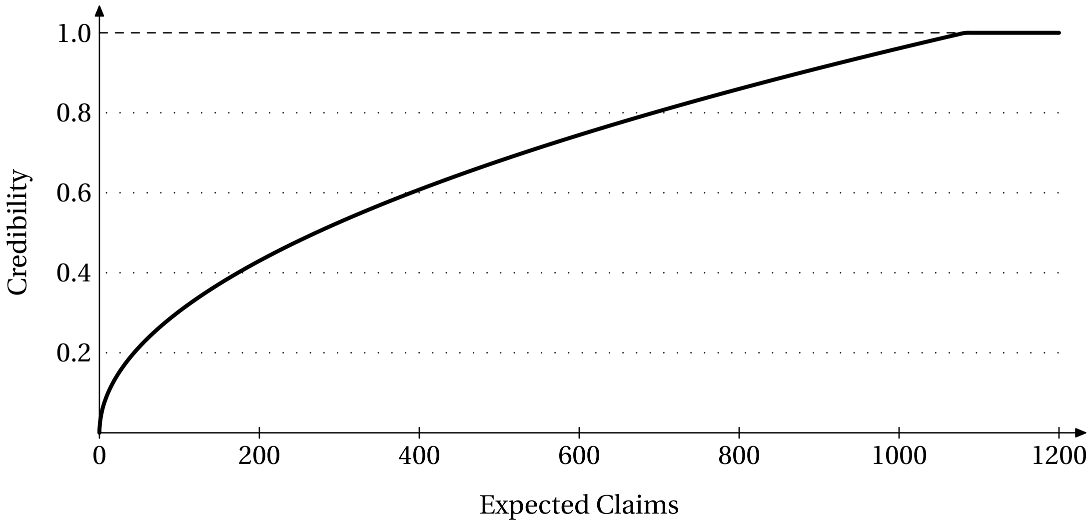

Figure 3.1: Partial credibility if $n_F = 1082$

SOLUTION: The credibility factor is $\sqrt{n/n_F}$, with $n = 6000$ and $n_F = 19{,}544$, so $Z = \sqrt{\frac{6{,}000}{19{,}544}} = 0.55408$, and the estimate is:

$$P_C = 16{,}500{,}000 + \sqrt{\frac{6{,}000}{19{,}544}}(15{,}600{,}000 - 16{,}500{,}000) = \boxed{16{,}001{,}332}. \qquad \textbf{(D)}$$

□

Exercises

3.1. You are given the following:

- Number of claims follows a Poisson distribution.
- Classical credibility methods are used.
- The standard for credibility is set so that the actual aggregate losses are within 5% of expected losses 90% of the time.
- 605 expected claims are required for 50% credibility.

Determine the coefficient of variation for the claim size distribution.

A. Less than 1.50
B. At least 1.50, but less than 2.00
C. At least 2.00, but less than 2.50
D. At least 2.50, but less than 3.00
E. At least 3.00

Exercises continue on the next page . . .

3.2. **[4B-S92:6]** (1 point) You are given the following information for a group of insureds:

Prior estimate of expected total losses	20,000,000
Observed total losses	25,000,000
Observed number of claims	10,000
Required number of claims for full credibility	17,500

Using the methods of classical credibility, determine the estimate for the group's expected total losses based upon the latest observation.

A. Less than 21,000,000

B. At least 21,000,000, but less than 22,000,000

C. At least 22,000,000, but less than 23,000,000

D. At least 23,000,000, but less than 24,000,000

E. At least 24,000,000

3.3. **[4B-F93:20]** (2 points) You are given the following:

- P = Prior estimate of pure premium for a particular class of business.
- O = Observed pure premium during latest experience period for same class of business.
- R = Revised estimate of pure premium for same class following observations.
- F = Number of claims required for full credibility of pure premium.

Based on the methods of classical credibility, determine the number of claims used as the basis for determining R.

A. $F\left(\dfrac{R-P}{O-P}\right)$ B. $F\left(\dfrac{R-P}{O-P}\right)^2$ C. $\sqrt{F}\left(\dfrac{R-P}{O-P}\right)$ D. $\sqrt{F}\left(\dfrac{R-P}{O-P}\right)^2$ E. $F^2\left(\dfrac{R-P}{O-P}\right)$

3.4. **[4-F03:35, STAM Sample Question #27]** You are given:

- X_{partial} = pure premium calculated from partially credible data.
- $\mu = \mathbf{E}[X_{\text{partial}}]$
- Fluctuations are limited to $\pm k\mu$ of the mean with probability P.
- $Z =$ credibility factor

Which of the following is equal to P?

A. $\Pr(\mu - k\mu \le X_{\text{partial}} \le \mu + k\mu)$

B. $\Pr(Z\mu - k\mu \le ZX_{\text{partial}} \le Z\mu + k)$

C. $\Pr(Z\mu - \mu \le ZX_{\text{partial}} \le Z\mu + \mu)$

D. $\Pr(1 - k \le ZX_{\text{partial}} + (1-Z)\mu \le 1 + k)$

E. $\Pr(\mu - k\mu \le ZX_{\text{partial}} + (1-Z)\mu \le \mu + k\mu)$

Exercises continue on the next page . . .