INSTITUTE OF ACTUARIAL & QUANTITATIVE STUDIES

Subject: SRM 2

Chapter: Unit 2

Category: Practice question

1. CT6 September 2010 Q4

An office worker receives a random number of e-mails each day. The number of emails per day follows a Poisson distribution with unknown mean μ. Prior beliefs about μ are specified by a gamma distribution with mean 50 and standard deviation 15. The worker receives a total of 630 e-mails over a period of ten days.

Calculate the Bayesian estimate of μ under all or nothing loss. [7]

2. Subject CT6 September 2010 Question 5

The table below shows aggregate annual claim statistics for three risks over a period of seven years. Annual aggregate claims for risk i in year j are denoted by X_{ij}.

Risk, i	$\bar{X}_i = \frac{1}{7}\sum_{j=1}^{7} X_{ij}$	$S_i^2 = \frac{1}{6}\sum_{j=1}^{7}\left(X_{ij} - \bar{X}_i\right)^2$
$i = 1$	127.9	335.1
$i = 2$	88.9	65.1
$i = 3$	149.7	33.9

(i) Calculate the credibility premium of each risk under the assumptions of EBCT Model 1. [6]

(ii) Explain why the credibility factor is relatively high in this case. [2]
[Total 8]

3. CT6 April 2011 Q2

An insurance company has collected data for the number of claims arising from certain risks over the last 10 years. The number of claims in the jth year from the ith risk is denoted by X_{ij} for i = 1, 2, ..., n and j = 1, 2, ..., 10. The distribution of X_{ij} for j = 1 2, ..., 10 depends on an unknown parameter θi and given θi the X_{ij} are independent identically distributed random variables.

(i) Give a brief interpretation of $E[s^2(\theta)]$ and $V[m(\theta))]$ under the assumptions of Empirical Bayes Credibility Theory Model 1. [2]

(ii) Explain how the value of the credibility factor Z depends on $E[s^2(\theta)]$ and $V[m(\theta)]$. [3] [Total 5]

4. CT6 April 2011 Q3

Let $y_1, \ldots, y_n$ be samples from a uniform distribution on the interval [0, θ] where θ > 0 is an unknown constant. Prior beliefs about θ are given by a distribution with density

$$f(\theta) = \{\alpha\beta^{\alpha}\theta^{-(1+\alpha)} \qquad \theta > \beta \; 0 \qquad \textit{otherwise}$$

UNIT 2

PRACTICE QUESTION

where α and β are positive constants.

(i) Show that the posterior distribution of θ given y_1 is of the same form as the prior distribution, specifying the parameters involved. [4]

(ii) Write down the posterior distribution of θ given $y_1, \ldots, y_n$. [2]
[Total 6]

5. CT6 September 2011 Q2

An accountant is using a psychic octopus to predict the outcome of tosses of a fair coin. He claims that the octopus has a probability $p > 0.5$ of successfully predicting the outcome of any given coin toss. His actuarial colleague is extremely sceptical and summarises his prior beliefs about p as follows: there is an 80% chance that $p = 0.5$ and a 20% chance that p is uniformly distributed on the interval [0.5,1]. The octopus successfully predicts the results of 7 out of 8 coin tosses.

Calculate the posterior probability that $p = 0.5$. [4]

6. CT6 September 2011 Q11

Five years ago, an insurance company began to issue insurance policies covering medical expenses for dogs. The insurance company classifies dogs into three risk categories: large pedigree (category 1), small pedigree (category 2) and non-pedigree (category 3). The number of claims n_{ij} in the ith category in the jth year is assumed to have a Poisson distribution with unknown parameter θ_i.

Data on the number of claims in each category over the last 5 years is set out as follows:

	Year						
	1	2	3	4	5	$\sum_{j=1}^{5} n_{ij}$	$\sum_{j=1}^{5} n_{ij}^2$
Category 1	30	43	49	58	60	240	12144
Category 2	37	49	58	52	64	260	13934
Category 3	26	31	18	37	32	144	4354

Prior beliefs about θ_1 are given by a gamma distribution with mean 50 and variance 25.

(i) Find the Bayes estimate of θ_1 under quadratic loss. [5]

UNIT 2

PRACTICE QUESTION

(ii) Calculate the expected claims for year 6 of each category under the assumptions of Empirical Bayes Credibility Theory Model 1 [6]

(iii) Explain the main differences between the approach in (i) and that in (ii). [3]

(iv) Explain why the assumption of a Poisson distribution with a constant parameter may not be appropriate and describe how each approach might be generalised. [3]
[Total 17]

7. Subject CT6 April 2012 Question 5

The total claim amount per annum on a particular insurance policy follows a normal distribution with unknown mean θ and variance 200^2 . Prior beliefs about θ are described by a normal distribution with mean 600 and variance 50^2 . Claim amounts x_1, x_2 .....x_n are observed over n years.

(i) State the posterior distribution of θ. [2]

(ii) Show that the mean of the posterior distribution of q can be written in the form of a credibility estimate. [3]

Now suppose that n = 5 and that total claims over the five years were 3,400.

(iii) Calculate the posterior probability that q is greater than 600. [2]
[Total 7]

8. CT6 April 2012 Q6

A proportion p of packets of a rather dull breakfast cereal contain an exciting toy (independently from packet to packet). An actuary has been persuaded by his children to begin buying packets of this cereal. His prior beliefs about p before opening any packets are given by a uniform distribution on the interval [0,1]. It turns out the first toy is found in the n_1 th packet of cereal.

(i) Specify the posterior distribution of p after the first toy is found. [3]

A further toy was found after opening another n_2 packets, another toy after opening another n_3 packets and so on until the fifth toy was found after opening a grand total of $n_1 + n_2 + n_3 + n_4 + n_5$ packets.

(ii) Specify the posterior distribution of p after the fifth toy is found. [2]

(iii) Show the Bayes' estimate of p under quadratic loss is not the same as the maximum likelihood estimate and comment on this result. [5]

UNIT 2

PRACTICE QUESTION

[Total 10]

9. CT6 September 2012 Q8

An insurer classifies the buildings it insures into one of three types. For Type 1 buildings, the number of claims per building per year follows a Poisson distribution with parameter l . Data are available for the last five years as follows:

Year	1	2	3	4	5
Number of type 1 buildings covered	89	112	153	178	165
Number of claims	15	23	29	41	50

(i) Determine the maximum likelihood estimate of l based on the data above. [5]

The insurer also has data for the other two types of building for all five years.
Define:
P_{ij} = the number of buildings insured in the j th year from type i and
Y_{ij} = the corresponding number of claims.

The five years of data can be summarised as follows:

Type (i)	$\bar{P}_i = \sum_{j=1}^{5} P_{ij}$	$\bar{X}_i = \sum_{j=1}^{5} \frac{Y_{ij}}{\bar{P}_i}$	$\sum_{j=1}^{5} P_{ij}\left(\frac{Y_{ij}}{P_{ij}} - \bar{X}_i\right)^2$	$\sum_{j=1}^{5} P_{ij}\left(\frac{Y_{ij}}{P_{ij}} - \bar{X}\right)^2$
Type 1	697	0.226686	1.527016	2.502737
Type 2	295	0.237288	0.96605	1.178133
Type 3	515	0.330097	4.53253	6.775614

$$\bar{X} = \sum_{i=1}^{3} \sum_{j=1}^{5} \frac{Y_{ij}}{\bar{P}} = 0.264101 \text{ where } \bar{P} = \sum_{i=1}^{3} \bar{P}_i$$

There are 191 buildings of Type 1 to be insured in year six.

(ii) Estimate the number of claims from Type 1 buildings in year six using Empirical Bayes Credibility Theory model 2. [6]

(iii) Explain the main differences between the approaches in parts (i) and (ii). [2]
[Total 13]

UNIT 2

PRACTICE QUESTION

10. CT6 April 2013 Q3

An actuary has a tendency to be late for work. If he gets up late then he arrives at work X minutes late where X is exponentially distributed with mean 15. If he gets up on time then he arrives at work Y minutes late where Y is uniformly distributed on [0,25]. The office manager believes that the actuary gets up late one third of the time.

Calculate the posterior probability that the actuary did in fact get up late given that he arrives more than 20 minutes late at work. [5]

11. Subject CT6 April 2013 Question 10

An insurance company has a portfolio of building insurance policies. The company classifies buildings into three types and believes that the number of claims on buildings of each type follows a Poisson distribution with parameters as shown:

Type	*Parameter*
1	λ
2	2λ
3	5λ

where l is an unknown positive constant. Actual claim numbers over the last five years have been as follows.

Here X_{ij} represents the number of claims from the i th type in the j th year:

Number of claims X_{ij}

Year (j)

Type (i)	5	4	3	2	1	$\sum_{j=1}^{5}(X_{ij}-\bar{X}_i)^2$
1	23	17	9	21	12	139.2
2	56	39	44	29	35	417.2
3	87	115	141	92	84	2,322.8

(i) Derive the maximum likelihood estimate of l . [5]

(ii) Estimate the average number of claims per year for each type of building using EBCT Model 1. [7]

(iii) Comment on the results of parts (i) and (ii). [2]

(iv) Explain the main weakness of the model in part (ii). [1]

UNIT 2

PRACTICE QUESTION

[Total 15]

12.CT6 September 2013 Q10

The number of service requests received by an IT engineer on any given day follows a Poisson distribution with mean μ. Prior beliefs about μ follow a gamma distribution with parameters α and λ. Over a period of *n* days the actual numbers of service requests received are 1 2 , , .

(i) Derive the posterior distribution of μ . [3]

(ii) Show that the Bayes estimate of μ under quadratic loss can be written as a credibility estimate and state the credibility factor. [2]

Now suppose that α =10 , λ = 2 and that the IT worker receives 42 requests in 6 days.

(iii) Calculate the Bayes estimate of μ under quadratic loss. [1]

Three quarters of requests can be resolved by telling users to restart their machine, and the time taken to do so follows a Pareto distribution with density:

$$f(x) = \frac{3 \times 20^3}{(20+x)^4} \quad \text{for } x > 0$$

One quarter of requests are much harder to resolve, and the time taken to resolve these follows a Weibull distribution with density:

$$f(x) = 0.4 \times 0.5 x^{-0.5} e^{-0.4 x^{0.5}} \quad \text{for } x > 0$$

(iv) (a) Calculate the probability that a randomly chosen request takes more than 30 minutes to resolve.
(b) Calculate the average time spent on each request.
(c) Calculate the expected total amount of time the IT worker spends dealing with service requests each day, using the estimate of m from part (iii). [5]

The IT worker's line manager is carefully considering his staffing requirements. He decides to model the time taken on each request approximately using an exponential distribution.

(v) (a) Fit an exponential distribution to the time taken per request using the method of moments.
(b) Calculate the probability that a randomly chosen request takes more than 30 minutes to resolve using this approximation.
(c) Comment briefly on your answer to part (v)(b). [2]

The IT engineer needs to devote more of his time to a separate project, so his firm have hired an assistant to help him. The assistant is just as fast at dealing with the straightforward requests, and the time taken to resolve these still follows the Pareto distribution given above. He is significantly slower at dealing with the difficult requests, and the time taken to resolve these now follows a Weibull distribution with density:

$$f(x) = c \times 0.5x^{-0.5}e^{-cx^{0.5}} \quad \text{for } x > 0$$

where c is a positive parameter. The line manager again fits an exponential distribution as an approximation to the time taken to service each request using the method of moments. His approximation results in an estimate that the probability that a random service request takes longer than 30 minutes to resolve is 10%.

(vi) Determine the value of c . [4]
[Total 17]

13. CT6 September 2014 Q6

For three years an insurance company has insured buildings in three different towns against the risk of fire damage. Aggregate claims in the *j*th year from the *i*th town are denoted by X_{ij} for i = 1, 2, 3 and j = 1, 2, 3. The data is given in the table below.

Town i	*Year j* 1	2	3
1	8,130	9,210	8,870
2	7,420	6,980	8,130
3	9,070	8,550	7,730

Calculate the expected claims from each town for the next year using the assumptions of Empirical Bayes Credibility Theory model 1. [10]

14. CT6 April 2015 Q5

An insurance company has for five years insured three different types of risk. The number of policies in the *j*th year for the *i*th type of risk is denoted by P_{ij} for i = 1, 2, 3 and j = 1, 2, 3, 4, 5. The average claim size per policy over all five years for the *i*th type of risk is denoted by $\bar{X}_i$. The values of P_{ij} and $\bar{X}_i$ are tabulated below.

	Number of policies					*Mean claim size*
Risk type i	*Year 1*	*Year 2*	*Year 3*	*Year 4*	*Year 5*	$\bar{X}_i$
1	17	23	21	29	35	850
2	42	51	60	55	37	720
3	43	31	62	98	107	900

The insurance company will be insuring 30 policies of type 1 next year and has calculated the aggregate expected claims to be 25,200 using the assumptions of Empirical Bayes Credibility Theory Model 2.

UNIT 2

PRACTICE QUESTION

Calculate the expected annual claims next year for risks 2 and 3 assuming the number of policies will be 40 and 110 respectively. [9]

15. Subject CT6 April 2015 Question 9

Let p be an unknown parameter and let f(p |x) be the probability density of the posterior distribution of p given information x .

(i) Show that under all-or-nothing loss the Bayes estimate of p is the mode of f(p|x) . [2]

John is setting up an insurance company to insure luxury yachts. In year 1 he will insure 100 yachts and in year 2 he will insure 100 + g yachts where g is an integer.

If there is a claim the insurance company pays a fixed sum of $1m per claim.
The probability of a claim on a policy in a given year is p . You may assume that the probability of more than one claim on a policy in any given year is zero. Prior beliefs about p are described by a Beta distribution with parameters a = 2 and b = 8 . Page 80

In year 1 total claims are $13m and in year 2 they are $20m.

(ii) Derive the posterior distribution of p in terms of g . [4]

(iii) Show that it is not possible in this case for the Bayes estimate of p to be the same under quadratic loss and all-or-nothing loss. [6]
[Total 12]

16. Subject CT6 October 2015 Question 4

A small island is holding a vote on independence. Two recent survey results are shown below:

Poll	*Sample size*	*Support for independence*
A	10	5
B	20	11

You should assume that the samples are independent. A politician is using a suitable uniform distribution as the prior distribution in order to estimate the proportion θ in favour of independence.

(i) Calculate an estimate of θ under the quadratic loss function. [3]

A rival politician decides to use instead a beta distribution as the prior, with parameters α and β , where α = β .

(ii) Determine the new estimate of θ under the 'all-or-nothing' loss function in terms of α . [4]
[Total 7]

UNIT 2

PRACTICE QUESTION

17. Subject CT6 October 2015 Question 5

Claims X each year from a portfolio of insurance policies are normally distributed with mean θ and variance τ^2 . Prior information is that θ is normally distributed with known mean μ and known variance σ^2 .

Aggregate claims over the last n years have been x_i for i = 1 to n , and you should assume that these are independent.

(i) Derive the posterior distribution of θ . [5]

(ii) Write down the Bayesian estimate of θ under quadratic loss. [1]

(iii) Show that the estimate in your answer to part (ii) can be expressed in the form of a credibility estimate, including statement of the credibility factor Z . [2]
[Total 8]

18. Subject CT6 October 2015 Question 7

A shipping insurance company has insured ships for six years, and classifies the ships it insures into three types.
Let:
P_{ij} be the number of ships insured in the jth year from type i ,
Y_{ij} be the corresponding number of claims.

The six years of data are summarised as follows:

Type (i)	$\bar{P}_i = \sum_{j=1}^{6} P_{ij}$	$\bar{X}_i = \sum_{j=1}^{6} \frac{Y_{ij}}{\bar{P}_i}$	$\sum_{j=1}^{6} P_{ij}\left(\frac{Y_{ij}}{P_{ij}} - \bar{X}_i\right)^2$	$\sum_{j=1}^{6} P_{ij}\left(\frac{Y_{ij}}{P_{ij}} - \bar{X}\right)^2$
Type 1	648	0.524 691	30.966 692	64.392 683
Type 2	981	0.145 062	4.689 264	42.240 804
Type 3	636	0.370 370	62.449 512	66.467 182

$$\bar{X} = \sum_{i=1}^{3} \sum_{j=1}^{6} \frac{Y_{ij}}{\bar{P}} = 0.297572, \text{ where } \bar{P} = \sum_{i=1}^{3} \bar{P}_i$$

There are 100 ships of Type 3 to be insured in year seven.

(i) Estimate the number of claims from Type 3 ships in year seven using empirical Bayes credibility theory (EBCT) Model 2. [6]

UNIT 2

PRACTICE QUESTION

The insurance company's actuary is considering using EBCT Model 1 instead.

(ii) Explain an advantage and a disadvantage of using EBCT Model 1 rather than EBCT Model 2. [2]
[Total 8]

19. Subject CT6 April 2016 Question 3

A child playing a game believes that a six sided die is unfair, and that he has a probability p > 1/6 of predicting the outcome of any given throw. His mother is less sure, and her prior beliefs about p are as follows:

· a 1/3 chance that p = 2/6 and

· a 2/3 chance that p = 1/6

The child accurately predicts the results of 4 out of 10 dice throws.

Calculate the posterior probability that p = 1/6 . [6]

20.CT6 September 2016 Q3

The table below shows aggregate annual claim statistics for four risks over a period of six years. Annual aggregate claims for risk *i* in year *j* are denoted by *Xij*.

$Risk, i$	$\overline{X_i} = \frac{1}{6}\sum_{j=1}^{6} X_{ij}$	$S_i^2 = \frac{1}{5}\sum_{j=1}^{6}\left(X_{ij} - \overline{X_i}\right)^2$
$i=1$	46.8	1227.4
$i=2$	30.2	1161.4
$i=3$	74.5	1340.3
$i=4$	60.7	1414.7

(i) Calculate the credibility premium of each risk under the assumptions of Empirical Bayes Credibility Theory (EBCT) Model 1. [7]

(ii) Comment on why the credibility factor is relatively low in this case. [2]
[Total 9]

21. Subject CT6 April 2017 Question 4

The number of claims on a portfolio of insurance policies in a given year follows a Poisson distribution with unknown mean λ . Prior beliefs about λ are specified by a gamma distribution with mean 60 and variance 360. Over a period of three and one-third years, the total number of claims is 200.

(i) Calculate the Bayesian estimate of l under all-or-nothing loss. [7]

UNIT 2

PRACTICE QUESTION

(ii) Comment on your result for part (i). [1]
[Total 8]

22. CT6 April 2017 Q7

An actuary is assessing three different insurance companies, A, B and C.
Corresponding claim amounts and number of policies are shown in the data below.

	Company A		*Company B*		*Company C*	
	\$m	*Policies*	*\$m*	*Policies*	*\$m*	*Policies*
2013	1.16	85	0.85	68	1.48	110
2014	1.18	88	1.02	82	1.52	132
2015	1.14	85	0.96	70	1.78	143
2016	1.32	92	0.87	80	1.92	165
Total	**4.8**	**350**	**3.7**	**300**	**6.7**	**550**

Company C has 180 policies to insure in 2017.

(i) Calculate its expected claim amount, using the assumptions underlying Empirical Bayes Credibility Theory (EBCT) Model 2. [11]

(ii) Discuss why it might be preferable to use EBCT Model 2 rather than EBCT Model 1 for this purpose. [2]
[Total 13]

23. Subject CT6 September 2017 Question 11

(i) Explain what is meant by a conjugate prior distribution. [1]

The random variables X_1X_2 , ,, X_n are independent and have density function:

$$P(X = x) = p(1-p)^x, 0 < p < 1$$

(ii) Show that the conjugate prior for p is a beta distribution. [3]

Assume that we have an independent sample X_1X_2 , ,, X_n from a geometric distribution with parameter p, with the prior density function for p given by:

$$f(p) = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} p^{\alpha-1}(1-p)^{\beta-1}, 0 < p < 1$$

(iii) Show that $E\left(\frac{1-p}{p}\right) = \frac{\beta}{(\alpha-1)}$. [3]

UNIT 2

PRACTICE QUESTION

(iv) Show that the posterior mean of this distribution can be expressed as a weighted average of the prior mean and the sample average, including statement of the credibility factor Z. [2]

Every day Amit and Bonnie catch the bus home from work at a bus stop next to their office. Most buses which arrive at the bus stop do not go to their destination. Denote the average number of buses they have to wait for as N such that the $(N + 1)^{th}$ bus to arrive at the bus stop goes to their destination. Amit's prior belief is that N = 10. Bonnie's prior belief is that N = 5. They both use a beta prior distribution with a = 5 but with different b

(v) Calculate the number of bus trips home required such that the absolute difference in Amit and Bonnie's posterior estimates for N is less than 0.5. [5]
[Total 14]

24. CT6 April 2018 Q3

An insurance company has collected data on the number of claims arising from certain risks over the last *n* years. The number of claims from the *ith* risk in the *jth* year is denoted by *Xij* for *i* = 1, 2, ..., *N* and *j* = 1, 2, ..., *n*.

The distribution of *Xij* depends on an unknown parameter Θ*i*. The *Xij* are independent identically distributed random variables given Θ*i*.

(i) Describe briefly what is meant by each of the following: $m(\theta)$, $s^2(\theta)$, $E(s^2(\theta))$, $\text{var}(m(\theta))$, and Z, when using Empirical Bayes Credibility Theory (EBCT) Model 1. [5]

(ii) Explain how the value of Z depends on the following factors: n, $E(s^2(\theta))$, $\text{var}(m(\theta))$. [5]
[Total 10]

25. CT6 September 2018 Q3

(i) State the fundamental difference between Bayesian estimation and Classical estimation. [2]

(ii) State three different loss functions which may be used under Bayesian estimation, indicating for each its link to the posterior distribution. [3]

The proportion, Θ, of the population of a particular country who use online banking is being estimated. Of a sample of 500 people, 326 do use online banking.

An actuary is estimating Θ using a suitable uniform distribution as a prior.

(iii) (a) Determine the posterior distribution of Θ.
(b) Calculate an estimate of Θ using the loss function that minimises the mean of the posterior distribution. [4]
[Total 9]

UNIT 2

PRACTICE QUESTION

26. CS1A April 2019 Q9

Consider a sample of 1,000 motor insurance policies. We assume that the annual total claim amounts per policy are independent and identically distributed. We denote by X the number of policies with a total amount of over £5,000 claimed in a calendar year, and assume that X has a Binomial distribution, X ~ Bin(p, 1,000), with expectation E[X] = 1,000p.

An analyst wishes to estimate the unknown proportion p of claims with amount greater than £5,000 per year.
(i) Derive the maximum likelihood estimator for p. [3]

Suppose now that the analyst has some prior knowledge about p and assumes a Beta prior distribution with density function $f(p) = \dfrac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} p^{\alpha-1}(1-p)^{\beta-1}$.

(ii) Derive the density of the Bayesian posterior distribution of p in terms of n, X, α and β. [2]

(iii) State the type of the posterior distribution of p with its parameters. [2]

(iv) Comment on the relationship between the prior distribution and the posterior distribution of p in this context. [2]

Assume that 50 policies out of 1,000 policies in an actual sample have a total claim amount of over £5,000.

(v) Estimate p using the MLE in (i). [1]

(vi) Estimate p using the Bayesian estimator under quadratic loss, based on the posterior distribution derived in parts (ii) and (iii). Assume that the parameters of the prior distribution are α = 2 and β = 2. [3]
(vii) Comment on the difference between the values estimated in (v) and (vi). [1]

(viii) State the Bayesian estimator from part (vi) in the form of a credibility interval, determining the credibility factor. [3] [Total 17]

27. CS1A September 2019 Q3

The table below shows the annual aggregate claim statistics for three risks over four years. The annual aggregate claim for risk i, in year j, is denoted by Xij.

UNIT 2

PRACTICE QUESTION

Risk i	$\bar{X}_i = \frac{1}{4}\sum_{j=1}^{4} X_{ij}$	$s_i^2 = \frac{1}{3}\sum_{j=1}^{4}\left(X_{ij} - \bar{X}_i\right)^2$
1	2,109	3,959,980
2	6,152	7,543,626
3	3,016	3,151,286

(i) Calculate the value of the credibility factor for Empirical Bayes Model 1. [4]

(ii) Comment on how each of the following features of the data affects the value of the credibility factor calculated in part (i):
(a) the number of years of data
(b) the variance of the claim amounts. [2]
[Total 6]

28. CS1A September 2019 Q7
An actuary has designed a new product to insure luxury apartments. If there is a claim, her insurance company pays a fixed sum of £1 million per claim. The probability of a claim on a policy in a given year is θ and the probability of more than one claim on a policy in any given year is zero.
The actuary's prior beliefs about θ are given by a Beta distribution with parameters a = 3 and b = 5.

In the first year, the company insured 300 apartments and in the second year it insured 300 + x apartments, where x is an integer. In year 1 the total amount of claims was
£39 million, while in year 2 it was £60 million.
(i) Show that the posterior distribution of θ is Beta with parameters 102 and 506 + x. [7]

(ii) Derive the Bayesian estimate of θ in terms of x, under quadratic loss. [2]

(iii) Derive the Bayesian estimate of θ in terms of x, under all-or-nothing loss. [4]

(iv) Justify that, in this case, the Bayesian estimate of θ cannot be the same under quadratic and all-or-nothing loss. [2]
[Total 15]

29. CS1A September 2020 Q8
A statistician has recorded the number of advertising telephone calls that their office received over 2 years. The statistician has recorded data Xij, which is the number of calls received in the ith quarter of the jth year (where i = 1, 2, 3, 4 and j = 1, 2):

	X_{i1}	X_{i2}	$\bar{X}_i$	$\Sigma_j (X_{ij} - \bar{X}_i)^2$
$i=1$	43	29	36	98
$i=2$	38	42	40	8
$i=3$	22	18	20	8
$i=4$	68	56	62	72

(i) Calculate values for:

(a) E[m(θ)]

(b) E[s^2(θ)]

(c) Var[m(θ)] [4]

(ii) Calculate an estimate for X_{13}, the number of advertising telephone calls that the statistician's office expects to receive in the first quarter of year 3, using your answers to part (i) and the assumptions of the Empirical Bayes Credibility Theory Model 1 (EBCT Model 1). [2]

(iii) (a) State two key assumptions underlying the EBCT Model 1.
(b) Explain what these assumptions mean for the data Xij above. [4]
[Total 10]

30. CS1A April 2021 Q9

The number of claims received by a motor insurance company on any given day follows a Poisson distribution with mean u. Prior beliefs about u are expressed through a gamma distribution with parameters a and b. Over a period of n days the observed number of claims received per day are x1, x2, ..., xn.

(i) Identify which one of the following is the posterior density of u:

A $f(u|x) \propto u^{b + \Sigma x_i - 1} e^{-(a+n)u}$

B $f(u|x) \propto u^{a + \Sigma x_i} e^{-(b+n+1)u}$

C $f(u|x) \propto u^{a + \Sigma x_i - 1} e^{-(b+n)u}$

D $f(u|x) \propto u^{b + \Sigma x_i + 1} e^{-(a+n-1)u}$

[3]

(ii) Write down the posterior density of the parameter u and specify its parameters. [2]

(iii) (a) Determine the Bayesian estimate of u under quadratic loss. [2]
(b) Write down the Bayesian estimate of u under quadratic loss as a credibility estimate and state the credibility factor. [2]

Suppose that a = 9, b = 3 and that the company receives 320 claims in total during a 6-day period.

UNIT 2

PRACTICE QUESTION

(iv) Calculate the Bayesian estimate of u under quadratic loss. [2]

(v) Calculate the variance of the posterior distribution of u. [2]

An industry expert suggests that prior beliefs about u are better expressed through a gamma distribution with parameters a = 18 and b = 6.

(vi) Explain how these prior beliefs would affect the variance of the posterior distribution of u, without explicitly calculating the variance of the posterior distribution. [2]
[Total 15]

31. CS1A September 2021 Q4

The number of pizzas ordered in a restaurant each day follows a Poisson distribution with unknown mean m. The prior distribution for m follows a gamma distribution with mean 35 and standard deviation 5. The restaurant receives 135 pizza orders over 7 days.

(i) Write down an expression of the prior probability density function for mleaving out any coefficient of proportionality. [3]

(ii) Identify which one of the following expressions gives the correct posterior probability density function for m.

A $f_{\text{posterior}}(m) \propto m^{135} e^{-7m}$

B $f_{\text{posterior}}(m) \propto m^{183} e^{-7.7m}$

C $f_{\text{posterior}}(m) \propto m^{184} e^{-8.4m}$

D $f_{\text{posterior}}(m) \propto m^{183} e^{-8.4m}$

[3]

(iii) Calculate a point estimate for the number of pizzas ordered each day, using Bayesian estimation under all-or-nothing loss. [4]

(iv) Calculate a point estimate for the number of pizzas ordered each day, using Bayesian estimation under squared-error loss. [2]
[Total 12]

32. CS1A September 2021 Q10

Total yearly aggregate claims in a particular company are modelled as a random variable X, where X is assumed to follow a Normal distribution with unknown mean μ and variance σ^2 = 12,000ˆ2.
Aggregate claims from the last 5 years are as follows:

UNIT 2

PRACTICE QUESTION

146,000   142,000   153,000   127,000   132,000

An analyst wishes to estimate the unknown parameter μ.

(i) Identify which one of the following gives the correct expression of the derivative of the log-likelihood function:

A $\frac{dl(\mu)}{d\mu} = -\sum_{i=1}^{n}(x_i - \mu)$

B $\frac{dl(\mu)}{d\mu} = \sum_{i=1}^{n}(x_i - \mu)$

C $\frac{dl(\mu)}{d\mu} = \frac{1}{\sigma^2}\sum_{i=1}^{n}(x_i - \mu)$

D $\frac{dl(\mu)}{d\mu} = -\frac{1}{\sigma^2}\sum_{i=1}^{n}(x_i - \mu)$

(ii) Calculate the maximum likelihood estimate for μ, using your answer to part (i). [1]

(iii) Calculate a 95% confidence interval for μ. [4]

The analyst assumes a Normal prior distribution for μ with density function

$$f(\mu) \propto e^{-\frac{(\mu-\mu_0)^2}{2\sigma_0^2}}, \quad \mu_0 > 0 \text{ and } \sigma_0 > 0.$$

For such a prior, the analyst derives the posterior distribution for μ as

$$p(\mu|\underline{x}) \propto \exp\left(-\frac{1}{2}(n\tau + \tau_0)\left(\mu - \frac{n\tau\bar{x} + \tau_0\mu_0}{n\tau + \tau_0}\right)^2\right)$$

where $\tau = \frac{1}{\sigma^2}$ and $\tau_0 = \frac{1}{\sigma_0^2}$.

Prior information about μ suggests that $\mu_0 = 150{,}000$ and $\sigma_0^2 = 10{,}204.08^2$.

(iv) Write down the distribution corresponding to the density p(μ|$\underline{x}$) above, with all its parameters values. [2]

(v) Comment on the relationship between the prior distribution and the posterior distribution of μ. [1]

(vi) Calculate the value of the Bayesian credibility estimate for μ under quadratic loss. [2]

(vii) Calculate an approximate 95% Bayesian interval for μ, based on its posterior distribution. [2]

(viii) Comment on the intervals estimated in parts (iii) and (vii). [1]

Another analyst assumes a Uniform prior distribution for μ with mean μ0 = 150,000 and variance σ_0^2 = (10,204.08)^2

(ix) Identify which one of the following gives the correct expression of the posterior distribution for μ:

UNIT 2

PRACTICE QUESTION

A $p(\mu|\underline{x}) \propto \left(\frac{\mu-\mu_0}{\sigma_0^2}\right) \exp\left(-\frac{n}{2\sigma^2}(\mu-\bar{x})^2\right)$

B $p(\mu|\underline{x}) \propto \exp\left(-\frac{n}{2\sigma^2}(\mu-\bar{x})^2\right)$

C $p(\mu|\underline{x}) \propto \exp\left(-\frac{1}{2}\left(\frac{n}{\sigma^2}+\frac{1}{\sigma_0^2}\right)\left(\mu-\frac{\frac{n\bar{x}}{\sigma^2}+\frac{\mu_0}{\sigma_0^2}}{\frac{n}{\sigma^2}+\frac{1}{\sigma_0^2}}\right)^2\right)$

D $p(\mu|\underline{x}) \propto (\mu-\mu_0)^2 \exp\left(-\frac{n}{2\sigma^2}(\mu-\bar{x})^2\right)$

[3] [Total 18]

33. CS1A April 2022 Q8

The time, T, until the next lorry arrives at a customs checkpoint at the border of a country is modelled with an exponential distribution, that is, T ~ Exp(λ), where λ is an unknown parameter. Time is measured in minutes.

(i) Identify which one of the following expressions gives the correct likelihood function L(λ) for the parameter λ, based on a sample of observed times until the next lorry arrives, ti,, i = 1, …, n:

A $L(\lambda|T) = \lambda^n \exp(-\lambda\sum t_i)$

B $L(\lambda|T) = \lambda^{n-1} \exp(-\lambda\sum t_i)$

C $L(\lambda|T) = \lambda^{n+1} \exp(-\lambda\sum t_i)$

D $L(\lambda|T) = \lambda \exp(-\lambda\sum t_i)$

[1]

An analyst uses Bayesian inference to obtain an estimate for λ. They choose a gamma distribution with parameters α and β as the prior distribution for λ.

(ii) Verify that the posterior distribution of the parameter λ is a gamma distribution with parameters α + n and β + $\sum t_i$. [4]

Assume that a total of 20 lorries have arrived at the checkpoint.

(iii) Determine the Bayesian estimator for λ, in terms of the parameters α and β, under quadratic loss based on this sample. [2]

(iv) Explain how to determine the Bayesian estimator for λ under all-or-nothing loss based on this sample. [3]

(v) Identify which one of the following options gives the correct Bayesian estimator for λ under all-or-nothing loss based on the sample given:

A $\lambda = \frac{\alpha}{\beta + 60}$

B $\lambda = \frac{\alpha + 19}{\beta + 60}$

C $\lambda = \frac{\alpha + 18}{\beta + 60}$

D $\lambda = \frac{\alpha + 20}{\beta + 60}$

[2]

(vi) Comment on the difference between the two estimators in parts (iii) and (v). [1]

[Total 13]

34. CS1A September 2022 Q6

Let x1, x2, ..., xn independent observations from a Bernoulli distribution with P(Xi = 1) = p where i = 1, ..., n. The parameter p has a beta prior distribution with parameters (a, b).

(i) Determine the posterior distribution of parameter p. [6]

(ii) Determine the Bayesian estimate of parameter p under quadratic loss. [1]

(iii) Determine the Bayesian estimate of parameter p under quadratic loss as a credibility estimate, stating the credibility factor. [2]

[Total 9]

35. CS1A April 2023 Q5

Two people are playing a game together, involving the toss of a single coin. The coin used is biased so that the probability of throwing a head is an unknown constant, h. It is known that h must be either 0.35 or 0.85. Prior beliefs about h are given by the following distribution:

P(h = 0.35) = 0.7 and P(h = 0.85) = 0.3

The coin is tossed 15 times, and nine heads are observed.

Determine the posterior probabilities for the two possible values of h. [7]

36. CS1A September 2023 Q7

Total losses in a particular company are modelled by a random variable Y with density function:

UNIT 2

PRACTICE QUESTION

$$f(y) = \begin{cases} \frac{c}{y^{c+1}}, & y > 1, \quad c > 0 \\ 0, & \text{otherwise.} \end{cases}$$

An analyst wishes to estimate the unknown parameter c.

(i) Identify which one of the following expressions gives the maximum likelihood estimate for parameter c:

A $\hat{c} = \frac{c}{\sum_1^n \log(y_i)}$

B $\hat{c} = \frac{cn}{\sum_1^n \log(y_i)}$

C $\hat{c} = \frac{n}{\sum_1^n \log(y_i)}$

D $\hat{c} = \frac{n+1}{\sum_1^n \log(y_i)}$. [3]

The analyst assumes a gamma prior distribution for c with parameters (a, b).

(ii) Determine the posterior distribution of c with all its parameters. [6]

(iii) Comment on the relationship between the prior distribution and the posterior distribution of c. [1]

(iv) Determine the Bayesian estimate of parameter c under quadratic loss. [2]

[Total 12]

37. CS1A September 2023 Q8

The time spent (in minutes) queuing in a line by customers at a bank is modelled by the random variable X with density:

$$f(x) = \theta x e^{\frac{-\theta x^2}{2}}, \; x > 0.$$

The parameter θ is assumed to follow a prior gamma distribution with parameters a and b.

(i) Identify which one of the following expressions is proportional to the posterior density of θ given a random sample $x1$, $x2$,...xn of observations from X:

UNIT 2

PRACTICE QUESTION

A $f(\theta|x) \propto \theta^{n+a} e^{-\theta\left(\frac{\sum_{i=1}^{n} x_i^2}{2}+b+1\right)}$

B $f(\theta|x) \propto \theta^{n+a-1} e^{-\theta\left(\frac{\sum_{i=1}^{n} x_i^2}{2}+b\right)}$

C $f(\theta|x) \propto \theta^{n+a-2} e^{-\theta\left(\frac{\sum_{i=1}^{n} x_i^2}{2}+b\right)}$

D $f(\theta|x) \propto \theta^{n+a-1} e^{-\theta\left(\frac{\sum_{i=1}^{n} x_i^2}{2}+b-1\right)}$. [3]

(ii) State the posterior distribution of θ in part (i) and its parameters. [3]

The waiting times of ten customers are recorded in the table below and the prior knowledge is determined with a = 4, b = 1.5.

2.5	1.25	3	1.5	5.5	4	3.25	2	1.5	1

(iii) Calculate a point estimate of the time spent queuing in the line using Bayesian estimation under all-or-nothing loss. [4]

[Total 10]