INSTITUTE OF ACTUARIAL & QUANTITATIVE STUDIES

Subject: P&S

Chapter: Unit 3

Category: Practice question

Central limit theorem, Sampling distributions, Theory of estimation 1

1. CT3 April 2013 Question 4

Consider a random sample, X_1, X_2 … X_n, from a normal $N(\mu, \sigma^2)$ distribution, with sample mean $\overline{X}$ and sample variance S^2.

i. Define carefully what it means to say that X_1, X_2 … X_n is a random sample from a normal distribution.

ii. State what is known about the distributions of $\overline{X}$ and S^2 in this case, including the dependencies between the two statistics.

iii. Define the t-distribution and explain its relationship with $\overline{X}$ and S^2.

2. CT3 April 2013 Question 7

A regulator wishes to inspect a sample of an insurer's claims. The insurer estimates that 10% of policies have had one claim in the last year and no policies had more than one claim. All policies are assumed to be independent.

i. Determine the number of policies that the regulator would expect to examine before finding 5 claims.

On inspecting the sample claims, the regulator finds that actual payments exceeded initial estimates by the following amounts:

£35 £120 £48 £200 £76

ii. Find the mean and variance of these extra amounts.

It is assumed that these amounts follow a gamma distribution with parameters α and λ.

iii. Estimate these parameters using the method of moments.

3. CT3 October 2013 Question 2

An insurance company experiences claims at a constant rate of 150 per year. Find the approximate probability that the company receives more than 90 claims in a period of six months.

4. CT3 October 2013 Question 3

The random variable X has a distribution with probability density function given by:

$$f(x) = \begin{cases} \dfrac{2x}{\theta^2} & ; \quad 0 \le x \le \theta \\ 0 & ; \quad x < 0 \text{ or } x > \theta \end{cases}$$

Unit 3

PRACTICE QUESTION

where θ is the parameter of the distribution.

i. Derive expressions in terms of θ for the expected value and the variance of X.

Suppose that X_1, X_2 … X_n is a random sample, with mean, X from the distribution of X.

ii. Show that the estimator $\hat{\theta} = \frac{3\bar{X}}{2}$ is an unbiased estimator of θ.

5. CT3 October 2013 Question 4

An actuary is considering statistical models for the observed number or claims, X, which occur in a year on a certain class of non-life policies. The actuary only considers policies on which claims do actually arise. Among the considered models is a model for which:

$$P(X=x) = -\frac{1}{\log(1-\theta)}\frac{\theta^x}{x}, \quad x=1, 2, 3, \ldots$$

where θ is a parameter such that $0 < \theta < 1$.

Suppose that the actuary has available a random sample X_1, X_2 … , X_n, with sample mean $\bar{X}$.

i. Show that the method of moments estimator (MME), $\tilde{\theta}$ satisfies the equation:

$$\bar{X}(1-\tilde{\theta})\log(1-\tilde{\theta}) + \tilde{\theta} = 0.$$

ii.

a. Show that the log likelihood of the data is given by:

$$l(\theta) \propto -n\log\{-\log(1-\theta)\} + \sum_{i=1}^{n} x_i \log(\theta).$$

b. Hence verify that the maximum likelihood estimator (MLE) of θ is the same as the MME.

iii. Suggest two ways in which the MLE of θ can be computed when a particular data set is given.

6. CT3 October 2013 Question 5

Consider a random sample consisting of the random variables X_1, X_2 … , X_n with mean μ and variance σ^2. The variables are independent of each other.

i. Show that the sample variance, S^2, is an unbiased estimator of the true variance σ^2.

Unit 3

PRACTICE QUESTION

Now consider in addition that the random sample comes from a normal distribution, in which case it is known that $\frac{(n-1)S^2}{\sigma^2} \sim \chi^2_{n-1}$

ii. (a) Derive the variance of S^2 in terms of σ and n.

(b) Comment on the quality of the estimator S^2 with respect to the sample size n.

7. CT3 April 2014 Question 8

Let $X_1, X_2 \dots, X_n$ be a random sample from a distribution with parameter θ and density function:

$$f(x) = \begin{cases} \frac{2x}{\theta^2} & ; \quad 0 \le x \le \theta \\ 0 & ; \quad x < 0 \text{ or } x > \theta \end{cases}$$

Suppose that $\underline{x} = (x_1, x_2, \dots, x_n)$ is a realisation of $X_1, X_2 \dots, X_n$.

i.

a. Derive the likelihood function $L(\theta; \underline{x})$ and produce a rough sketch of its graph.

b. Use the graph produced in part (i)(a) to explain why the maximum likelihood estimate of θ is given by $X_{(n)} = \max\{X_1, X_2 \dots, X_n\}$.

Let $X_{(n)} = \max\{X_1, X_2 \dots, X_n\}$ be the estimator of θ, that is the random variable corresponding to $x_{(n)}$.

ii.

a. Show that the cumulative distribution function of the estimator $X_{(n)}$ is given by:

$$F_{X_{(n)}}(x) = \left(\frac{x}{\theta}\right)^{2n}$$

for $0 \le x \le \theta$.

b. Hence, derive the probability density function of the estimator $X_{(n)}$.

c. Determine the expected value $E(X_{(n)})$ and the variance $V(X_{(n)})$.

d. Show that the estimator $\frac{2n+1}{2n} X_{(n)}$ is an unbiased estimator of θ.

iii.

a. Derive the mean square error of the estimator given in part (ii)(d).

b. Comment on the consistency of this estimator.

8. CT3 September 2014 Question 7

Unit 3

PRACTICE QUESTION

Consider the following discrete distribution with an unknown parameter p for the distribution of the number of policies with 0, 1,2, or more than 2 claims per year in a portfolio of n independent policies.

number of claims	0	1	2	more than 2
Probability	2p	p	0.25p	1-3.25p

We denote by X0 the number of policies with no claims, by X1 the number of policies with one claim and by X2 the number of policies with two claims per year. The random variable X = X0 + X1 + X2 is then the number of policies with at most two claims.

i. Derive an expression for the maximum likelihood estimator $\tilde{p}$ of parameter p in terms of X and n.

ii. Show that the estimator obtained in part (i) is unbiased.

The following frequencies are observed in a portfolio of n = 200 policies during the year 2012:

number of claims	0	1	2	more than2
observed frequency	123	58	13	6

A statistician proposes that the parameter p can be estimated by $\tilde{p}$= 58/200 = 0.29 since p is the probability that a randomly chosen policy leads to one claim per year.

iii. Estimate the parameter p using the estimator derived in part (i).

iv. Explain why your answer to part (iii) is different from the proposed estimated value of 0.29.

An alternative model is proposed where the probability function has the form:

number of claims	0	1	2	more than2
probability	P	2p	0.25p	1-3.25p

v. Explain how the maximum likelihood estimator suggested in part (i) needs to be adapted to estimate the parameter p in this new model.

vi. Suggest a suitable test to use to make a decision about which of the two models should be used based on empirical data.

9. CT3 April 2015 Question 6

Unit 3

PRACTICE QUESTION

Let $X_1, X_2, \ldots, X_6$ be a random sample from a population following a Gamma(2,1) distribution. Consider the following two estimators of the mean of this distribution:

$$\hat{\theta}_1 = \overline{X} \text{ and } \quad \hat{\theta}_2 = \frac{9}{30}(X_1 + X_2 + X_3) + \frac{1}{30}(X_4 + X_5 + X_6)$$

where $\overline{X}$ is the mean of the sample.

i. Determine the sampling distribution of $\overline{X}$ using moment generating functions.
ii. Derive the bias of each estimator $\hat{\theta}_1$ and $\hat{\theta}_2$
iii. Derive the mean square error of each estimator $\hat{\theta}_1$ and $\hat{\theta}_2$
iv. Compare the efficiency of the two estimators $\hat{\theta}_1$ and $\hat{\theta}_2$

10. CT3 October 2015 Question 10

The random variables $X_1, X_2, \ldots, X_n$ are independent from each other and all follow a Poisson distribution with parameter λ.

i. Derive the maximum likelihood estimator of λ based on $X_1, X_2, \ldots, X_n$. You are not required to verify that your answer corresponds to a maximum.

11. CT3 April 2016 Question 5

Players A and B play a game of "heads or tails", each throwing 50 fair coins. Player A will win the game if she throws 5 or more heads than B; otherwise, B wins. Let the random variables X_A and X_B denote the numbers of heads scored by each player and $D = X_A = X_B$.

i. Explain why the approximate asymptotic distribution of D is normal with mean 0 and variance 25.

ii. Determine the approximate probability that player A wins any particular game, based on your answer in part (i).

12. CT3 April 2016 Question 6

A statistician is sent a summary of some data. She is told that the sample mean is 9.46 and the sample variance is 25.05. She decides to fit a continuous uniform distribution to the data.

i. Estimate the parameters of the distribution using the method of moments.

The full data are sent later and are given below:

3.5 5.4 7.3 8.5 9.2 10.3 11.4 20.1

Unit 3

PRACTICE QUESTION

ii. Comment on the results in part (i) in the light of the full data.

13. CT3 April 2016 Question 7

A random sample is taken from an exponential distribution with parameter λ. The sample contains some censored observations for which we only know that the value is greater than 3. The observed values are given in the following table:

I	1	2	3	4	5	6	7	8	9	10
X_i	1.3	1.8	2.1	2.2	2.2	2.4	>3	>3	>3	>3

Estimate the parameter λ using the method of maximum likelihood. You are not required to verify that your answer corresponds to the maximum.

14. CT3 October 2016 Question 9

A statistical model is used to describe the total loss, S (in pounds), experienced in a certain portfolio of an insurance company over a period of one year. The total loss is given by:

$$S = X_1 + X_2 + \cdots + X_N$$

where X_i gives the size of the loss from claim i = 1,...,N.

N is a random variable representing the number of claims per year and follows a Poisson distribution. The X_i's are independent, identically distributed according to a gamma distribution with parameters α and λ, and are also independent of N.

Data from previous years show that the average number of claims per year was 14, while the average size of claims was £500 and their standard deviation was £150.

i. Estimate the parameters α and λ using the method of moments.

ii. Estimate the mean and the variance of the total loss S using the information from the data above.

Now suppose that the value of parameter α is known to be equal to α* and n = 5 claims have been made in a particular year with average size again £500.

iii.

a. Derive an expression for the maximum likelihood (ML) estimate of the parameter λ in terms of α*. You should verify that your answer corresponds to a maximum.
b. Derive the asymptotic distribution of the ML estimator of the parameter λ in terms of α*.
c. Comment on the validity of the distribution in part (iii)(b).

Unit 3

PRACTICE QUESTION

Now suppose that the values of both parameters α and λ are unknown and n claims have been made in a particular year.

iv.

a. Show that the ML estimate, $\hat{\alpha}$ of the parameter α needs to satisfy the equation:

$$\log(\hat{\alpha}) - \frac{\Gamma'(\hat{\alpha})}{\Gamma(\hat{\alpha})} = \log(\bar{x}) - \frac{\sum_{i=1}^{n} \log(x_i)}{n}$$

where Γ'($\hat{\alpha}$) denotes the first derivative of Γ'($\hat{\alpha}$) with respect to $\hat{\alpha}$

b. Comment on how the ML estimates of the parameters α and λ can be obtained in this case.

15. CT3 April 2017 Question 5

Let $X_1, X_2, \ldots, X_n$ be a sequence of independent, identically distributed random variables with finite mean μ and finite (non-zero) variance σ^2.

i. State the central limit theorem (CLT) in terms of the sum $\sum_{i=1}^{n} X_i$

Assume now that each X_i, $i = 1, 2, \ldots, 50$, follows an exponential distribution with parameter λ = 2 and let $Y = \sum_{i=1}^{50} X_i$

ii. Determine the approximate distribution of Y together with its parameters using the CLT.

iii. State the exact distribution of Y together with its parameters.

iv. Comment on the shape of the distribution of Y based on your answers to parts (ii) and (iii).

16. CT3 April 2017 Question 7

An investigation at a large airport focuses on the delay with which i flights arrive. The delay time X, in minutes, is the difference between the actual time of arrival and the scheduled arrival time of delayed flights. Assume that X has an exponential distribution with parameter $\lambda > 0$.

i. Derive the estimator $\hat{\lambda}$ for λ using the method of moments.

The following table shows the observed values of X for a random sample of ten delayed flights.

45	20	120	90	60	30	45	90	60	150

ii. Estimate the value of λ for this sample using the method of moments.

To gain further insight into the distribution of flight delays, it is suggested that the time at which a flight is scheduled to arrive during a day has an impact on the delay.

Unit 3

PRACTICE QUESTION

Therefore, assume now that X_i has an exponential distribution with a parameter λ that depends on the scheduled arrival time as follows:

$$X_i \sim Exp(\lambda_i) \text{ with } \lambda_i = \theta Z_i$$

where the random variable Z_i describes the scheduled arrival time (in minutes) after midnight on the day of arrival for the i^{th} randomly selected delayed flight and θ > 0 is a parameter in this model.

iii. Derive the maximum likelihood estimator $\hat{\theta}$ for the parameter θ. You should show that your solution is indeed a maximum.

17. CT3 April 2017 Question 8

An actuary models the number of claims X per year per policy as a discrete random variable with the following distribution:

Number of claims	0	1	2	3	More than 3
Probability	*	P	p/2	p/4	p/8

where p is an unknown parameter.

i. Show that $P[X=0] = \frac{8-15p}{8}$

ii. Determine the range of possible values of p.

In a sample of n independent policies there are N_0 policies with no claims during a year, N_1 policies with one claim, N_2 policies with two claims and N_3 policies with three claims. There are also some policies with more than three claims.

iii. Show that the maximum likelihood estimator $\hat{p}$ for p based on observations of N_0, … , N_3 in a sample of n independent claims is given by:

$$\hat{p} = \frac{8}{15}\frac{n - N_0}{n}$$

You do not need to check that your solution is a maximum.

iv. Explain why the distribution of N_0 is a Binomial distribution specifying its parameters.

v. Verify that $\hat{p}$ is an unbiased estimator for p.

Assume that in a sample of size n = 300 there were 100 policies with no claims during the previous year.

vi. Determine the value of the variance of the estimator $\hat{p}$.

Unit 3

PRACTICE QUESTION

The insurance company has now decided to limit the maximum number of claims per year to four per policy, but otherwise continue to use the distribution above. The claim amount of any individual claim is assumed to have a normal distribution with expectation 100 and standard deviation 20. Let S denote the total amount claimed in a portfolio of 300 independent policies during a year. We assume that claim amounts are independent of each other and independent of the number of claims.

Let X be the number of claims per policy per year and Y be the total number of claims per year.

vii.

(a) Show that E(X) = 3.25p and Var(X) = 7.25p − 10.5625p^2.

Assume now that p = 0.2.

(b) Determine E(Y) and Var(Y).

(c) Determine the expected value and the standard deviation of S.

18. CT3 September 2017 Question 8

The two random variables X_1 and X_2 are independent from each other and follow a uniform $U(-\theta, \theta)$ distribution, where $\theta > 0$ is a parameter.

Let $\hat{\theta}_1 = 3Z$ denote a possible estimator of θ, where $Z = \max(X_1, X_2)$.

i. Show that the probability density function of Z is given by $f_Z(z) = \frac{z+\theta}{2\theta^2}$ by first deriving its cumulative distribution function.

ii. Show that $E(Z) = \frac{\theta}{3}$

iii.

(a) Derive the bias of $\hat{\theta}_1$

(b) Derive an expression for the mean squared error (MSE) of $\hat{\theta}_1$ in terms of the unknown parameter θ.

Let $\hat{\theta}_2 = 2Z$ denote a different estimator of θ, where again Z = $\max(X_1, X_2)$.

iv.

(a) Show that bias $(\hat{\theta}_2) = \frac{-\theta}{3}$

(b) Show that MSE$(\hat{\theta}_2) = \theta^2$

Unit 3

PRACTICE QUESTION

v. Comment on how good the two estimators are, based on your answers in parts (iii) and (iv).

19. CS1 April 2019 Question 5

i. State the central limit theorem for independent identically distributed random variables X_1, X_2 … , X_n with finite mean μ and finite (non-zero) variance σ^2.

ii. Show that if the random variable B has the binomial distribution with parameters (n,p), then $\frac{B-np}{\sqrt{np(1-p)}}$ approximately follows a standard normal distribution for large n, using the central limit theorem.

Two players have played a large number of independent games. In a sample of 100 of these games, one player has won 57 games and the other player has won 43.

iii. Derive a 95% confidence interval for the probability p that the first player wins a given game, using the normal approximation in part (ii).

20. CS1 September 2019 Question 2

Let X_1, X_2, … , X_n be a random sample consisting of independent random variables with mean μ and variance σ^2. Consider the sample mean:

$$\bar{X} = \frac{\sum_{i=1}^{n} X_i}{n}$$

i. Derive the expected value of $\bar{X}$.

ii. Derive the variance of $\bar{X}$

iii. Comment on the variance of variable $\bar{X}$ as compared to the variance of X_i.

An actuary is interested in exploring the difference in the size of claim losses from two insurance portfolios, and can take samples of claims from these portfolios.

iv. Explain how the answer to part (iii) can affect the precision of the actuary's comparison.

21. CS1A April 2021 Q1

A random variable, X, is modeled using a gamma distribution with parameters α = 50 and λ = 0.25.

i. Calculate an approximate value for P(X > 270) using the chi-square distribution.

ii. Calculate an approximate value for P(X > 270) using the central limit theorem.

iii. Comment on the difference between your answers to parts (i) and (ii).

Unit 3

PRACTICE QUESTION

22. CS1A April 2021 Q4

Consider a random sample of size n = 25 from a Normal distribution with mean 10, variance 4 and sample variance S^2.

i. Write down the sampling distribution of S^2.

ii. Calculate, using your answer in part (i), the expected value of S^2.

iii. Calculate, using your answer in part (i), the variance of S^2.

23. CS1A April 2021 Q6

A tutor believes that the number of exams passed by students sitting three different exams follows a binomial distribution with parameters n = 3 and p. A random sample of 120 students showed the following results:

Number of exams passed	0	1	2	3
Number of students	40	60	15	5

(i) (a) Identify which one of the following corresponds to the log likelihood function of p given the observed data:

A $\log L \propto 255\log(1-p) + 105\log(p)$

B $\log L \propto 115\log(1-p) + 80\log(p)$

C $\log L \propto 265\log(1-p) + 115\log(p)$

D $\log L \propto 175\log(1-p) + 85\log(p)$

(b) Show, using your answer to part (i)(a), that the maximum likelihood estimate for p is $\hat{p}$ = 0.2917. You are not required to check that it is a maximum.

24. CS1A September 2021 Q1

A random sample of size 15 is taken from a Normal distribution with mean 19 and variance 2.

i. Write down the sampling distribution of S^2.

ii. Explain why your answer in part (i) is valid for this random sample.

25. 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 = $12{,}000^2$.

Aggregate claims from the last 5 years are as follows: 146,000, 142,000, 153,000, 127,000, 132,000

An analyst wishes to estimate the unknown parameter μ.

Unit 3

PRACTICE QUESTION

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).

26. CS1A April 2022 Q4

i. Describe what is meant by each of the following:

(a) A random sample

(b) A statistic.

A new political party is interested in the level of support it would have among the voters in a particular country. The random variable X is defined as:

$$X = \begin{cases} 1, & \text{if the voter would support the party,} \\ 0, & \text{otherwise.} \end{cases}$$

A random sample of 50 voters are presented with a simple summary of the party's policies and asked if they would support this new party. The random sample is represented by $X_1, X_2, \ldots, X_{50}$.

ii. (a) Identify a suitable population together with a possible parameter of interest.

(b) Determine, using your answer to part (ii)(a), the sampling distribution of the statistic:

$$Y = \sum_{i=1}^{50} X_i$$

27. CS1A APRIL 2022 Q5

Unit 3

PRACTICE QUESTION

Let $X_1, X_2, \ldots, X_n$ be independent identically distributed random variables following a Poisson(m) distribution. Suppose that, rather than observing the random variables precisely, only the events $X_i = 0$ or $X_i > 0$ are observed for i = 1, 2, …, n.

Let Y be a random variable with:

$$Y_i = \begin{cases} 0, & X_i = 0 \\ 1, & X_i > 0 \end{cases}$$

for i = 1, 2, …, n.

i. Explain why the distribution of Y_i is a Bernoulli (p) distribution with parameter $p = 1 - e^{-m}$.

ii. Identify which one of the following expressions gives the correct likelihood function based on observations $y_1, \ldots, y_n$ in terms of $\bar{y} = \frac{1}{n}\sum_{i=1}^{n} y_i$ and the unknown parameter m.

A $\quad L(m) = (1 + e^{-m})^{n\bar{y}}(e^{m})^{n-n\bar{y}}$

B $\quad L(m) = (1 - e^{m})^{n\bar{y}}(e^{-m})^{n-n\bar{y}}$

C $\quad L(m) = (1 - e^{-m})^{n\bar{y}}(e^{-m})^{n-n\bar{y}}$

D $\quad L(m) = (1 - e^{-m})^{n\bar{y}}(e^{-m})^{n+n\bar{y}}$

iii. Derive an expression for the Maximum Likelihood Estimate (MLE) $\hat{m}$ of m in terms of $\bar{y} = \frac{1}{n}\sum_{i=1}^{n} y_i$

iv. State the condition that $\hat{m}$ and L(m) must satisfy for $\hat{m}$ to maximise the likelihood function.

28. CS1A APRIL 2022 Q6

The size of claims on a certain type of motor insurance policy are modelled as a random variable X with Probability Density Function (PDF)

$$f(x; \alpha, \beta) = \alpha \frac{\beta^{\alpha}}{x^{\alpha+1}}, \qquad x \geq \beta, \quad \alpha, \beta > 0.$$

i. Identify which one of the following expressions gives the correct log likelihood function in terms of a random sample $(x_1, x_2, \ldots, x_n)$ and the unknown parameters α and β:

Unit 3

PRACTICE QUESTION

A $l(\alpha, \beta) = n\log\alpha + n\alpha\log\beta + (\alpha+1)\sum_{i=1}^{n}\log x_i$

B $l(\alpha, \beta) = \log\alpha + n\alpha\log\beta - (\alpha+1)\sum_{i=1}^{n}\log x_i$

C $l(\alpha, \beta) = n\log\alpha + n\log\beta - (\alpha+1)\sum_{i=1}^{n}\log x_i$

D $l(\alpha, \beta) = n\log\alpha + n\alpha\log\beta - (\alpha+1)\sum_{i=1}^{n}\log x_i$

ii. Derive the MLE $\hat{\alpha}$ of parameter α as a function of parameter β, for a random sample (x_1, x_2, …, x_n).

iii. Comment on the behaviour of the PDF of X when β increases.

iv. Determine the MLE $\hat{\beta}$ of parameter β based on your comment in part (iii).

The values (in $) of a sample of 10 claims are given in the table below:

x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8	x_9	x_{10}
10,000	12,000	8,000	16,000	20,000	19,000	17,000	22,000	18,000	5,000

v. Calculate the mean and standard deviation of the natural logarithm of the sample.

vi. Calculate the MLEs $\hat{\alpha}$ and $\hat{\beta}$ based on the sample.

Unit 3

PRACTICE QUESTION

Theory of estimation 2, Hypothesis testing 1

1. CT3 April 2011 Question 9

Claims on a certain type of policy are such that the claim amounts are approximately normally distributed.

i. A sample of 101 such claim amounts (in £) yield a sample mean of £416 and sample standard deviation of £72. For this type of policy:

a) Obtain a 95% confidence interval for the mean of the claim amounts.
b) Obtain a 95% confidence interval for the standard deviation of the claim amounts.

The company makes various alterations to its policy conditions and thinks that these changes may result in a change in the mean, but not the standard deviation, of the claim amounts. It wants to take a random sample of claims in order to estimate the new mean amount with a 95% confidence interval equal to:

sample mean $\pm$£10.

ii. Determine how large a sample must be taken, using the following as an estimate of the standard deviation:

a) The sample standard deviation from part (I).
b) The upper limit of the confidence interval for the standard deviation from part (i)(b).

iii. Comment briefly on your two answers in (ii)(a) and (ii)(b).

2. CT3 October 2011 Question 9

In a recent study of attitudes to a proposed new piece of consumer legislation ("proposal X") independent random samples of 200 men and 200 women were asked to state simply whether they were "for" (in favour of), or "against", the proposal.
The resulting frequencies, as reported by the consultants who carried out the survey, are given in the following table:

	Men	*Women*
For	138	130
Against	62	70

i. Carry out a formal chi-squared test to investigate whether or not an association exists between gender and attitude to proposal X.

Note: in this and any later such tests in this question you should state the P-value of the data and your conclusion clearly.

At a subsequent meeting to discuss these and other results, the consultants revealed that they had in fact stratified the survey, sampling 100 men and 100 women in England and 100 men and 100 women in Wales. The resulting frequencies were as follows:

	England		Wales	
	Men	*Women*	*Men*	*Women*
For	82	66	56	64
Against	18	34	44	36

A chi-squared test to investigate whether or not an association exists between gender and attitude to proposal X in England gives x2 = 6.653, while an equivalent test for Wales gives x2 = 1.333.

ii. a) Find the P-value for each of the chi-squared tests mentioned above and state your conclusions regarding possible association between gender and attitude to proposal X in England and in Wales.

b) Discuss the results of the survey for England and Wales separately and together, quoting relevant percentages to support your comments.

iii. A different survey of 200 people conducted in each of England, Wales, and Scotland gave the following percentages in favour of another proposal:

	England	*Wales*	*Scotland*
% in favour of proposal	62%	53%	58%

A chi-squared test of association between country and attitude to the proposal gives x2 = 3.332 on 2 degrees of freedom, with P-value 0.189.

Suppose a second survey of the same size is conducted in the three countries and results in the same percentages in favour of the proposal as in the first survey. The results of the two surveys are now combined, giving a survey based on the attitudes of 1,200 people.

a) State (or find) the results of a second chi-squared test for an association between country and attitude to the proposal, based on the overall survey of 1,200 people.

b) Comment briefly on the results.

3. CT3 April 2012 Question 12

Consider a random sample $X_1, \ldots, X_4$ of size k = 400 . Statistician A wants to use a x2 -test to test the hypothesis that the distribution of X_i is a binomial distribution with parameters n = 2 and unknown p based on the following observed frequencies of outcomes of X_i:

Possible realisation of X_i	0	1	2
Frequency	90	220	90

Unit 3

PRACTICE QUESTION

i. Estimate the parameter p using the method of moments.

ii. Test the hypothesis that X_i has a binomial distribution at the 0.05 significance level using the data in the above table and the estimate of p obtained in part (i).

Statistician B assumes that the data are from a binomial distribution and wants to test the hypothesis that the true parameter is p_0= 0.5.

iii. Explain whether there is any evidence against this hypothesis by using the estimate of p in part (i) and without performing any further calculations.

Statistician C wants to test the hypothesis that the random variables X_i have a binomial distribution with known parameters n= 2 and p = 0.5.

iv. Write down the null hypothesis and the alternative hypothesis for the test in this situation.

v. Carry out the test at the significance level of 0.05 stating your decision.

vi. Explain briefly the relationship between the test decisions in parts (ii), (iii) and (v), and in particular whether there is any contradiction.

4. CT3 October 2013 Question 6

A researcher obtains samples of 25 items from normally distributed measurements from each of two factories. The sample variances are 2.86 and 9.21 respectively.

i. Perform a test to determine if the true variances are the same.

ii. For each factory calculate central 95% confidence intervals for the true variances of the measurements.

iii. Comment on how your answers in parts (i) and (ii) relate to each other.

5. CT3 April 2014 Question 6

In an opinion poll, a sample of 100 people from a large town were asked which candidate they would vote for in a forthcoming national election with the following results:

Candidate	A	B	C
Supporters	32	47	21

Unit 3

PRACTICE QUESTION

i. Determine the approximate probability that candidate B will get more than 50% of the vote.

A second opinion poll of 150 people was conducted in a different town with the following results:

Candidate	A	B	C
Supporters	57	56	37

ii. Use an appropriate test to decide whether the two towns have significantly different voting intentions.

6. CT3 September 2014 Question 6

In a medical study conducted to test the suggestion that daily exercise has the effect of lowering blood pressure, a sample of eight patients with high blood pressure was selected. Their blood pressure was measured initially and then again, a month later after they had participated in an exercise programme.

The results are shown in the table below:

Patient	*1*	*2*	*3*	*4*	*5*	*6*	*7*	*8*
Before	155	152	146	153	146	160	139	148
After	145	147	123	137	141	142	140	138

i. Explain why a standard two-sample t-test would not be appropriate in this investigation to test the suggestion that daily exercise has the effect of lowering blood pressure.

ii. Perform a suitable t-test for this medical study. You should clearly state the null and alternative hypotheses.

7. CT3 April 2015 Question 9

An insurance company has calculated premiums assuming that the average claim size per claim for a certain class of insurance policies does not exceed £20,000 per annum. An actuary analyses 25 such claims that have been randomly selected. She finds that the average claim size in the sample is £21,000 and the sample standard deviation is £2,500. Assume that the size of a single claim is normally distributed with unknown expectation α and variance σ^2.

i. Calculate a 95% confidence interval for α based on the sample of 25 claims.

ii. Perform a test for the null hypothesis that the expected claim size is not greater than £20,000 at a 5% significance level.

Unit 3

PRACTICE QUESTION

iii. Discuss whether your answers to parts (i) and (ii) are consistent.

iv. Calculate the largest expected claim size, α_0, for which the hypothesis $\alpha \leq \alpha_0$ can be rejected at a 5% significance level based on the sample of 25 claims.

The insurer is also concerned about the number of claims made each year. It is found that the average number of claims per policy was 0.5 during the year 2011. When the analysis was repeated in 2012 it was found that the average number of claims per policy had increased to 0.6. These averages were calculated on the basis of random samples of 100 policies in each of the two years.

Assume that the number of claims per policy per year has a Poisson distribution with unknown expectation λ and is independent from the number of claims in any other year or for any other policy.

v. Perform a test at 5% significance level for the null hypothesis that $\lambda = 0.6$ during the year 2011.

vi. Perform a test to decide whether the average number of claims has increased from 2011 to 2012.

8. CT3 October 2015 Question 9

A survey team is using satellite technology to measure the height of a mountain. This is an established technology and the variability of measurements is known. On each satellite pass over a mountain, they get a measurement that they know lies within ±5m of the true height with a 95% probability. The survey height is given by the mean of the measurements. They assume that all measurements are independent and follow a normal distribution with mean equal to the true height.

i. a) Show that the standard deviation of a single measurement is 2.551m.

b) Determine how many satellites passes over a mountain are required to have a 95% confidence interval for the true height with width less than 1m.

In a mountain range there are two summits which appear to have a similar height. The team manages to get 20 measurements for each summit and finds there is a difference of 1.6m between the mean survey height of the two summits.

ii. Perform a statistical test of the null hypothesis that the summits' true heights are the same, against the alternative that they are different.

At the same time the team is testing a new system on these two summits. They again get 20 measurements on each summit with an estimated standard deviation on the first summit of 2.5m and on the second of 2.6m. This system also measures the difference in survey heights between the two peaks to be 1.6m.

iii. Perform a statistical test of the same hypotheses as in part (ii) when heights are measured by the new system, justifying any assumptions you make.

iv. Comment on your answers to parts (ii) and (iii).

9. CT3 April 2016 Question 10

Consider a large portfolio of insurance policies and denote the claim size (in £) per claim by X. A random sample of policies with a total of 20 claims is taken from this portfolio and the claims made for these policies are reported in the following table:

Claim i	1	2	3	4	5	6	7	8	9	10
Claim size x_i	130	164	170	173	173	175	177	183	183	184

Claim i	11	12	13	14	15	16	17	18	19	20
Claim size x_i	185	186	197	202	208	213	215	229	233	272

For these data: $\sum x_i = 3{,}852$ and $\sum x_i^2 = 759{,}348$.

i. Calculate the mean, the median and the standard deviation of the claim size per claim in this sample.

ii. Determine a 95% confidence interval for the expected value E[X] based on the above random sample, stating any assumptions you make.

iii. Determine a 95% confidence interval for the standard deviation of X based on the above random sample.

iv. Explain briefly why the confidence interval in part (iii) is not symmetric around the estimated value of the standard deviation.

An actuary assumes that the number of claims from each policy has a Poisson distribution with an unknown parameter λ. In a new sample of 50 policies the actuary has observed a total of 80 claims yielding an estimated value of $\hat{\lambda}=1.6$ for the parameter λ.

v. Determine a 95% confidence interval for λ using a normal approximation.

vi. Determine the smallest required sample size n for which a 95% confidence interval for λ has a width of less than 0.5. You should use the same normal approximation as in part (v),

Unit 3

PRACTICE QUESTION

and assume that the estimated value of λ will not change.

Now assume that the true value of λ is 1.6 and the values calculated in part (i) are the true values. Also assume that all claims in the portfolio are independent and the claim sizes are independent of the number of claims.

vii. Determine the expected value and the standard deviation of the total amount of all claims from a portfolio of 5,000 insurance policies.

10. CT3 October 2016 Question 4

Consider two portfolios, A and B, of insurance policies and denote by X_A the number of claims received in portfolio A and by X_B the number of claims received in portfolio B during a calendar year. The observed numbers of claims received during the last calendar year are 134 for portfolio A and 91 for portfolio B. X_A and X_B are assumed to be independent and to have Poisson distributions with unknown parameters β_A and β_B.

Determine an approximate 99% confidence interval for the difference $\beta_A - \beta_B$. You may use an appropriate normal distribution.

11. CT3 April 2017 Question 6

We consider the impact that different types of cars have on the amount spent on fuel per month. Three different types of cars are considered: small, medium and large. For each type of car a group of 15 drivers are asked about the amount of money (in £) spent on fuel per month. The results are summarised in the following table

Type of car	*Small*	*Medium*	*Large*
Sample mean	70	75	83
Sample standard deviation	16	19	16

For example, the 15 drivers of medium sized cars spent on average £75 per month with a sample standard deviation of £19.

i. Perform a one-way analysis of variance to test the hypothesis that the type of car has no impact on the monthly amount spent on fuel. For some further investigation, only the difference between small and large cars is considered.

ii. Determine a 95% confidence interval for the difference between the average amount spent on fuel for small cars and large cars, stating any assumptions you make.

iii. Test the null hypothesis that the average fuel costs for small and large cars are the same at a 5% significance level against the alternative that the fuel costs for small and large cars are different.

12. CT3 April 2017 Question 9

A statistician is examining the survey methodology of a country's national statistics department. It conducts much of its data collection by telephoning individuals selected at random and asking them questions.

i. Comment on whether this methodology will give a random sample.

ii. Comment on whether this methodology will give a representative random sample of the population.

13. CT3 October 2017 Question 5

In an election between two candidates A and B in a large district, a sample poll of 100 voters chosen at random, indicated that 55% were in favour of candidate A.

i. Calculate a 95% confidence interval for the proportion of all voters in favour of candidate A based on the above sample.

A candidate is elected if they win more than 50% of the votes. We want a test in which the alternative hypothesis is that support for candidate A is such that she will win the election.

ii. a) Write down the hypotheses for this test in terms of a suitable parameter.

b) Explain whether or not the confidence interval in part (i) can be used to test the hypothesis in part (ii)(a) at the 5% level of significance.

It has been reported in the news that a new poll estimates support for candidate A at 52%, with a margin of error of no more than ±2% with confidence 95%.

iii. Determine the minimum size of the sample of voters that was taken in this new poll.

14. CT3 April 2018 Question 10

A large pension scheme regularly investigates the lifestyle of its pensioners using surveys. In successive surveys it draws a random sample from all pensioners in the scheme and it obtains the following data on whether the pensioners smoke.

Survey 1: Of 124 pensioners surveyed, 36 were classed as smokers.

Survey 2: Of 136 pensioners surveyed, 25 were classed as smokers.

An actuary wants to investigate, using statistical testing at a 5% significance level, whether there have been significant changes in the proportion of pensioners, p, who smoke in the entire pension scheme.

i. Perform a statistical test, without using a contingency table, to determine if the proportion p has changed from the first survey to the second.

When a third survey is performed it is found that 26 out of the 141 surveyed pensioners, are smokers.

ii. Perform a statistical test using a contingency table to determine if the proportion p is different among the three surveys.

iii. a) Calculate the proportion of smokers in the third survey.
b) Comment on your answers to parts (i) and (ii).

15. CS1A April 2019 Q3

The number of claims on a certain type of policy follows a Poisson distribution with claim rate l per year. For a group of 200 independent policies of this type, the total number of claims during the last calendar year was 82.

Determine an approximate 95% confidence interval for the true annual claim rate for this type of policy based on last year's claims.

16. CS1A September 2019 Q9

An actuary wants to model a particular type of claim size and has been advised to use a Gamma distribution with probability distribution function:

$$f(x, \alpha, \theta) = \frac{x^{\alpha-1}}{\Gamma(\alpha)\theta^{\alpha}} e^{-\frac{x}{\theta}}, \quad 0 < x < \infty, \quad \alpha > 0, \quad \theta > 0.$$

i. Show, using moment generating functions, that:

(a) $E(X) = \alpha\theta$

(b) $E(X^2) = \alpha(\alpha+1)\theta^2$

(c) $E(X^3) = \alpha(\alpha+1)(\alpha+2)\theta^3$.

The shape parameter alpha is assumed to be $\alpha = 4$.

Unit 3

PRACTICE QUESTION

ii. (a) Determine the variance of the claim size distribution in terms of θ.

(b) Calculate the coefficient of skewness of the claim size distribution, which is defined as:

$$\frac{E\left[(X-E(X))^3\right]}{\left\{E\left[(X-E(X))^2\right]\right\}^{1.5}}$$

Let $X_1, X_2, \ldots, X_n$ be a random sample of n claim sizes for such claims.

iii. Show that the maximum likelihood estimator (MLE) of θ is given by:

$$\hat{\theta} = \frac{\bar{X}}{4}$$

iv. Show that $\hat{\theta}$ is an unbiased estimator of θ.

A sample of n = 100 claim sizes yields

$\sum x_i = 796.2$ and $\sum x_i^2 = 8{,}189.4$.

v. Calculate the MLE of θ.

vi. (a) Calculate the sample variance.

(b) Compare the result in part (vi)(a) with the variance of the distribution evaluated at $\hat{\theta}$.

The sample coefficient of skewness is given as 1.12.

vii. Comment on its comparison with the coefficient of skewness of the distribution, calculated in part (ii)(b).

viii Calculate an appropriate 95% confidence interval for q by using an approximate 95% confidence interval for the mean of the distribution of the claim size.

ix. (a) Determine the variance of the distribution of q at both lower and upper limits of the confidence interval calculated in part (viii).

(b) Comment on the result in part (ix)(a) with reference to your answer in part (vi)(a) above.

17. CS1A September 2020 Q9

For an empirical investigation into the amount of rent paid by tenants in a town, data on income X and rent Y have been collected. Data for a total of 300 tenants of one bedroom flats have been recorded. Assume that X and Y are both Normally distributed with expectations μ_X and μ_Y, and variances σ_X^2 and σ_Y^2. S_X and S_Y are the sample standard deviation for random samples of X and Y, respectively.

The random variable Z_X is defined as

$$Z_X = 299\frac{S_X^2}{\sigma_X^2}.$$

i. State the distribution of Z_X and all of its parameters.

ii. Write down the expectation and variance of Z_X.

iii. Explain why the distribution of Z_X is approximately Normal.

iv. Calculate values of an approximate 2.5% quantile and 97.5% quantile of the distribution of Z_X using your answers to parts (ii) and (iii).

In the collected sample, the mean income is \$1,838 with a realised sample standard deviation of \$211, the mean rent is \$608 with a realised sample standard deviation of \$275 and $\Sigma x_i y_i = 348 \times 10^6$.

v. Calculate a 95% confidence interval for the mean income.

vi. Calculate a 95% confidence interval for the mean rent.

vii. Calculate an approximate 95% confidence interval for the variance of income using your answer to part (iv).

18. CS1A April 2021 Q7

A telecommunications company has performed a small empirical study comparing phone usage in rural and urban areas, collecting data from a total of 35 people who use their phones independently. The average number of hours that each person spent using their phone during a week is denoted by Y.

In the following table, $\overline{Y}$, denotes the sample mean of Y in rural and urban areas, and S_Y denotes the sample standard deviations; that is

Unit 3

PRACTICE QUESTION

$S_Y^2 = \frac{1}{n-1}\sum_{i=1}^{n}(Y_i - \bar{Y})^2$.

	Sample size n	$\bar{Y}$	S_Y
Rural areas	15	3.7	2.1
Urban areas	20	4.4	1.9

A statistical test is to be performed, at the 5% significance level, to determine whether the null hypothesis that mean phone usage in rural areas is the same as mean phone usage in urban areas, i.e. for:

H_0: phone usage is equal versus
H_1: phone usage is not equal.

i. State a suitable distribution for the test statistic with its parameter(s).

ii Justify any assumption(s) required to perform this test.

iii. Identify which one of the following options gives the correct value of the test statistic for this test:

A –1.031
B –0.519
C –3.019
D –1.455.

iv. Write down the conclusion of the test including the relevant critical value(s) from the Actuarial Formulae and Tables.

v. Determine a 95% confidence interval for the mean phone usage (hours per week) for rural areas, stating any assumption(s) you make.

19. CS1A September 2021 Q5

The probability that a claim is made on a car insurance policy in a particular year is 0.06. The policies are assumed to be independent among them. 500 of these policies are selected at random.

i. Calculate the probability that no more than 40 of these policies will result in a claim during the year, stating any approximations you make.

Unit 3

PRACTICE QUESTION

Past data from the insurer indicate that the standard deviation of claim amounts is £75. The insurer wishes to construct a 95% confidence interval for the mean claim amount, with an interval width of £10.

ii. Calculate the sample size needed to achieve this level of accuracy for a 95% confidence interval.

20. 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 $\sigma^2 = 12{,}000^2$.

Aggregate claims from the last 5 years are as follows:

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

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

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

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

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

iii. Calculate a 95% confidence interval for μ.

Unit 3

PRACTICE QUESTION