INSTITUTE OF ACTUARIAL & QUANTITATIVE STUDIES

Subject: P&S 2

Chapter: Unit 1

Category: Practice question

1. CT3 April 2013 Question 4

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

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

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

iii. Define the t-distribution and explain its relationship with $\bar{X}$ 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:

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

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 $X_1, X_2 \ldots X_n$ 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 yearon 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 $X_1, X_2 \ldots, X_n$, with sample mean $\bar{X}$.

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

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

ii.

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

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

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 $X_1, X_2 \dots, X_n$ 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.

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:

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

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.

Unit 1

PRACTICE QUESTION

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

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 X_0 the number of policies with no claims, by X_1 the number of policies with one claim and by X_2 the number of policies with two claims per year. The random variable $X = X_0 + X_1 + X_2$ 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.

Unit 1

PRACTICE QUESTION

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 beadapted 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 shouldbe used based on empirical data.

9. CT3 April 2015 Question 6

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:

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

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

i. Determine the sampling distribution of $\bar{X}$ using moment generating functions.
ii. Derive the bias of each estimator $\widehat{\theta_1}$ and $\widehat{\theta_2}$
iii. Derive the mean square error of each estimator $\widehat{\theta_1}$ and $\widehat{\theta_2}$
iv. Compare the efficiency of the two estimators $\widehat{\theta_1}$ and $\widehat{\theta_2}$

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 ifshe 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 variance25.

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

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 interms 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).

Now suppose that the values of both parameters α and λ are unknown and n claims have been made ina 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 $\sum_{i=1}^{n} X_i$

Unit 1

PRACTICE QUESTION

Assume now that each Xi, i = 1, 2, ..., 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.

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

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

where the random variable Z_i describes the scheduled arrival time (in minutes) after midnight on theday 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 thefollowing distribution:

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

Unit 1

PRACTICE QUESTION

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 duringthe previous year.

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

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 individualclaim 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).

Unit 1

PRACTICE QUESTION

(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 $\widehat{\theta_1} = 3Z$ 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 $\widehat{\theta_1}$

(b) Derive an expression for the mean squared error (MSE) of $\widehat{\theta_1}$ in terms of the unknown parameter θ.

Let $\widehat{\theta_2} = 2Z$ denote a different estimator of θ, where again $Z = \max(X_1, X_2)$.

iv.

(a) Show that bias $(\widehat{\theta_2}) = \frac{-\theta}{3}$

(b) Show that MSE$(\widehat{\theta_2}) = \theta^2$

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 $X_1, X_2 \ldots , X_n$ 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 $X_1, X_2, \ldots, X_n$ be a random sample consisting of independent random variables with mean μ andvariance 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 modelled using a gamma distribution with parameters $\alpha = 50$ and $\lambda = 0.25$.
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).

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

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

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

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 $i = 1, 2, \ldots, n$.

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:

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.