



**Subject: Probability and
Statistics II**

Chapter: Unit 1

Category: Practice Questions

1. IFoA CT3 October 2012 Question 6

A random sample of size n is taken from a gamma distribution with parameters $\alpha = 8$ and $\lambda = 1/\theta$. The sample mean is $\bar{X}$, and θ is to be estimated.

- i. Determine the method of moments estimator (MME) of θ .
- ii. Find the bias of the MME determined in part (i).
- iii. :
 - a) Determine the mean square error of the MME of θ .
 - b) Comment on the efficiency of the MME of θ based on your answer in part (iii) (a).

2. IFoA CT3 April 2013 Question 4

Consider a random sample, $X_1, X_2 \dots X_n$, 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 $X_1, X_2 \dots X_n$ 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 .

3. IFoA 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.

4. IFoA 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.

5. IFoA CT3 October 2013 Question 3

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

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

where θ is the parameter of the distribution.

- (i) Derive expressions in terms of θ for the expected value and the variance of $\bar{X}$.

Suppose that $X_1, X_2, \dots, X_n$ is a random sample, with mean $\bar{X}$, from the distribution of X .

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

6. IFoA CT3 October 2013 Question 4

An actuary is considering statistical models for the observed number of 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:

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

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

Suppose that the actuary has available a random sample $X_1, X_2, \dots, X_n$, 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.

7. IFoA 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 normal distribution, in which case it is known that $\frac{(n-1)S^2}{\sigma^2} \sim \chi_{n-1}^2$.

- ii. :
- Derive the variance of S^2 in terms of σ and n .
 - Comment on the quality of the estimator S^2 with respect to the sample size n .

8. IFoA 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:

$$f(x) = \frac{2x}{\theta^2} ; 0 \leq x \leq \theta$$

$$0 ; x < 0 \text{ or } x > \theta$$

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

- i. :
- Derive the likelihood function $L(\theta; \underline{x})$ and produce a rough sketch of its graph.
 - 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. :
- Show that the cumulative distribution function of the estimator $X_{(n)}$ is given by:

$$F_{X_{(n)}}(x) = \left(\frac{x}{\theta}\right)^{2n} \text{ for } 0 \leq x \leq \theta$$

- 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.

9. IFoA 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 than 2
observed frequency	123	58	13	6

A statistician proposes that the parameter p can be estimated by $p = 58/200 = 0.29$ 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 than 2
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.

10. IFoA CT3 April 2015 Question 6

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

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

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 $\hat{\theta}_1$ 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 $\hat{\theta}_2$.

11. IFoA CT3 October 2015 Question 10

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

- i. Derive the maximum likelihood estimator of λ based on $X_1, X_2, \dots, X_n$. You are not required to verify that your answer corresponds to a maximum.
- ii. Derive an expression for an approximate 95% confidence interval for λ under the situation in part (i), using the Cramer-Rao lower bound.

Suppose that instead of observing the values of $X_1, X_2, \dots, X_n$ precisely, we only observe that for K of these variables we have $X_i = 0$, while (for the remaining variables we have $X_i > 0$.

- iii. :
 - a. Derive the maximum likelihood estimator of λ when only this information is available. You are not required to verify that your answer corresponds to a maximum.
 - b. Explain why we need to observe at least one variable to be equal to zero for the estimator in part (iii) (a) to provide a sensible answer.

- iv. State, with reasons, whether you would prefer to use the estimator or of A in part (i) or that in part (iii).

12. IFoA 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 $D = X_A - X_B$.

- Explain why the approximate asymptotic distribution of D is normal with mean 0 and variance 25.
- Determine the approximate probability that player A wins any particular game, based on your answer in part (i).

13. IFoA 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.

- 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

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

14. IFoA 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.

15. IFoA 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 $i = 1, \dots, N$. 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).

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

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

16. IFoA CT3 April 2017 Question 5

Let $X_1, X_2, \dots, X_n$ 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 , $i = 1, 2, \dots, 50$, follows an exponential distribution with parameter $\lambda = 2$ 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).

17. IFoA 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:

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

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 $\theta > 0$ 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.

18. IFoA 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 $N_0, \dots, N_3$ in a sample of n independent claims is given by:

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

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 $\hat{p}$.

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.

19. IFoA 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 $MSE(\hat{\theta}_2) = \theta^2$.

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

20. IFoA CS1 April 2019 Question 5

i. State the central limit theorem for independent identically distributed random variables $X_1, X_2, \dots, X_n$ 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).

21. IFoA CS1 September 2019 Question 2

Let $X_1, X_2, \dots, X_n$ 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.