1. [$7 \times 1 = 7$ marks] **Circle the letter corresponding to the correct answer.**

(a) Consider an experiment in which data $y_1, y_2, ..., y_n$ are collected and the sample mean $\bar{y} = \frac{1}{n}\sum_{i=1}^{n} y_i$ is used to estimate the mean of the population. Which of the following statements is FALSE?

A: The variability of the sampling distribution of the sample mean is affected by the sample size n.

B: The variability of the sampling distribution of the sample mean is affected by the standard deviation of the population.

$\boxed{\text{C:}}$ The location of the sampling distribution of the sample mean is affected by the sample size n.

D: How often the sample mean is within one unit of the population mean is affected by the standard deviation of the population.

E: The skewness of the sampling distribution of the sample mean is affected by the skewness of the population.

(b) Which of the following statements is TRUE?

A: An estimator $\tilde{\theta}$ is a known quantity which can be used to estimate θ.

$\boxed{\text{B:}}$ An estimator $\tilde{\theta}$ is a random variable and its distribution is called the sampling distribution.

C: An estimate $\hat{\theta}$ is a random variable and its distribution is called the sampling distribution.

D: The true value of the parameter θ is known as soon as we have collected the data.

(c) The width of a 20% likelihood interval for θ for data from a $Binomial\,(n, \theta)$ distribution

$\boxed{\text{A:}}$ decreases as n increases.

B: increases as n increases.

C: does not change as n increases.

D: None of the above.

(d) A pivotal quantity for the unknown parameter θ is

A: a function of the observed data $y_1, y_2, ..., y_n$.

B: a function of $Y_1, Y_2, ..., Y_n$ whose distribution depends only on θ.

$\boxed{\text{C:}}$ a function of $Y_1, Y_2, ..., Y_n$ and θ whose distribution does not depend on θ.

D: None of the above.

(e) The width of a 95% confidence interval for μ for data from a $G(\mu, \sigma)$ distribution with known value of σ

A: does not change as n increases.

$\boxed{\text{B:}}$ decreases as n increases.

C: increases as n increases.

$\boxed{\text{D:}}$ increases as n decreases.

(f) Suppose it is reasonable to model observed data $y_1, y_2, ..., y_n$ using a *Exponential* (θ) distribution. Let thetahat $= \bar{y} = \frac{1}{n}\sum\limits_{i=1}^{n} y_i$. Which of the following functions in R should be used to calculate the relative likelihood function for θ?

A: ExpRLF <- function(x) {log((thetahat/x)^n*exp(n*(1-thetahat/x)))}

B: ExpRLF <- function(x) {log((x/thetahat)^n*exp(n*(1-x/thetahat)))}

$\boxed{\text{C:}}$ ExpRLF<- function(x) {(thetahat/x)^n*exp(n*(1-thetahat/x))}

D: ExpRLF <- function(x) {(x/thetahat)^n*exp(n*(1-x/thetahat))}

E: None of the above.

(g) Suppose it is reasonable to model observed data $y_1, y_2, ..., y_{25}$ using a *Exponential* (θ) distribution. Suppose also that the maximum likelihood estimate of θ is $\hat{\theta} = 2$ is observed. Which of the following statements in R provides the maximum likelihood estimate of $P(Y \leq 3)$ if $Y \backsim$ *Exponential* (θ)?

A: dexp(3,1/2)

$\boxed{\text{B:}}$ pexp(3,1/2)

C: qexp(3,1/2)

D: rexp(3,1/2)

E: None of the above.

2. [8 marks] Suppose $y_1, y_2, \ldots, y_n$ is an observed random sample from the distribution with probability density function

$$f(y;\theta) = \theta y^{\theta-1}, \qquad 0 < y < 1, \;\; \theta > 0.$$

(a) [4 marks] Derive the maximum likelihood estimate of θ based on the data $y_1, y_2, \ldots, y_n$. **Clearly show all your steps.**

The likelihood function is

$$L(\theta) = \prod_{i=1}^{n} f(y_i;\theta) = \prod_{i=1}^{n} \theta y_i^{\theta-1} = \theta^n \left(\prod_{i=1}^{n} y_i\right)^{\theta-1} \quad \text{for } \theta > 0$$

or more simply

$$L(\theta) = \theta^n \left(\prod_{i=1}^{n} y_i\right)^{\theta} \quad \theta > 0$$

The log likelihood is

$$l(\theta) = n\log\theta + \theta\log\left(\prod_{i=1}^{n} y_i\right) \quad \theta > 0$$

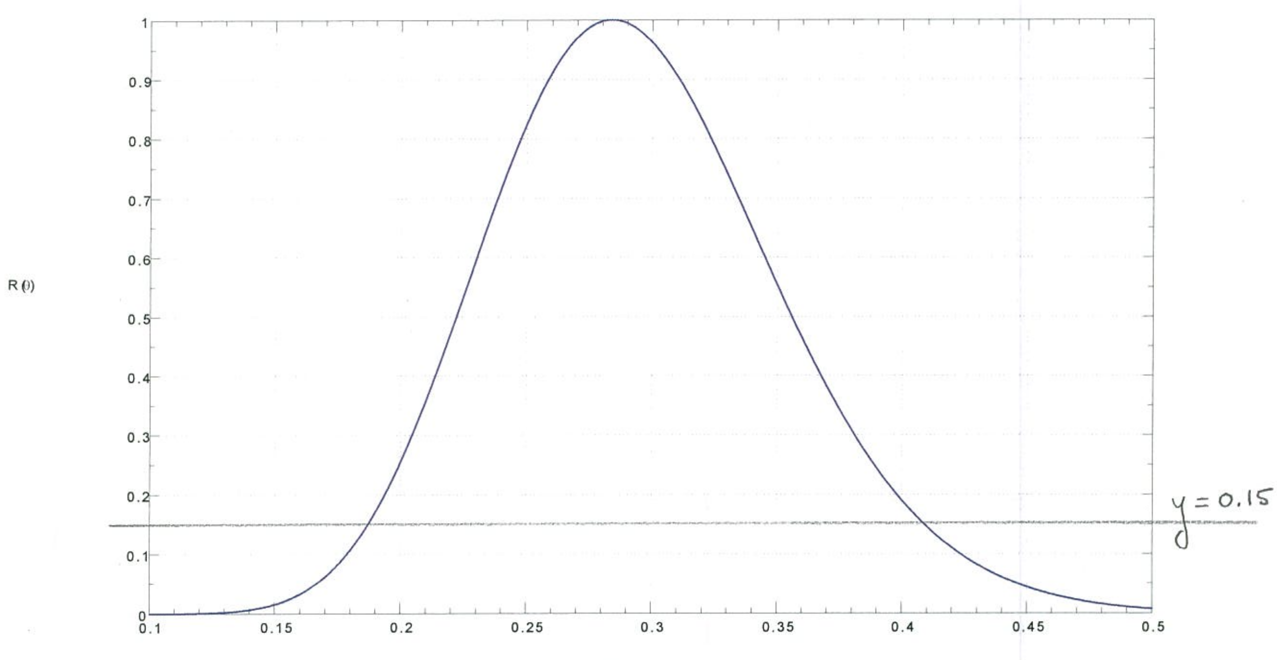
Solving

$$l'(\theta) = \frac{n}{\theta} + \log\left(\prod_{i=1}^{n} y_i\right) = \frac{n}{\theta} + \sum_{i=1}^{n}\log(y_i) = 0$$

gives the maximum likelihood estimate

$$\hat{\theta} = \frac{-n}{\log\left(\prod_{i=1}^{n} y_i\right)} = \frac{-n}{\sum_{i=1}^{n}\log(y_i)}$$

(b) [2 marks] The following is a plot of the relative likelihood function for θ for a given set of data:



From the plot, graphically determine a 15% likelihood interval for θ. Use two decimal places

[0.19 , 0.41]

(c) [2 marks] Complete the following R code to obtain the upper limit of the 15% likelihood interval for θ:

```
RLF <- function(x) {exp(n*log(x/thetahat)+n*(1-x/thetahat))}
uniroot(function(x) RLF(x)-0.15,lower= 0.35 , upper= 0.45 )
```

Note: Lower limit must be a number ≥ 0.3 but < 0.41. The upper limit must be a number > 0.41 but < 0.5.

3. [3 marks] In a study on the ability of rats to navigate a maze, a researcher records the length of time it took 21 different rats to find food at the end of the maze. Times in seconds are assumed to follow a $G(\mu, \sigma)$ distribution. A previous study has shown that $\sigma = 10.2$ seconds. Let y_i = time to complete the maze for the ith rat, $i = 1, 2, \ldots, 21$. The researcher's data gave $\sum_{i=1}^{21} y_i = 1068.4$ and $\sum_{i=1}^{21} y_i^2 = 56453.7$.

Write your final answer only in the space provided.

(a) [2 marks] A 90% confidence interval for μ is (use 3 decimal places)

[47.215 , 54.538]

$$\bar{y} \pm 1.645 \times \frac{\sigma}{\sqrt{n}} = \frac{1068.4}{21} \pm 1.645 \left(\frac{10.2}{\sqrt{21}}\right) = 50.87619 \pm 3.661478 = [47.21471, 54.53767]$$

(b) [1 mark] Circle the letter corresponding to the best interpretation of the interval found in (a).

A. The probability that the true value of μ is contained in the interval is 0.9.

$\boxed{\text{B.}}$ If the experiment is repeated many times, then we expect 90% of the intervals to contain the true value of μ.

C. We are 90% confident that $\mu = \hat{\mu}$.

D. All of the above.

4. [$7 \times 1 = 7$ marks] **Write your final answer only in the space provided.**

(a) **Without using Chi-squared tables** determine the following: (use 3 decimal places)

(i) If $X \backsim \chi^2(1)$ then $P(X > 2.25) =$ ___0.134___.

$$P(X \geq 2.25) = 2P\left(Z \geq \sqrt{2.25}\right) = 2\left[1 - P(Z \leq 1.5)\right] = 2(1 - 0.93319) = 0.13362$$

(ii) If $X \backsim \chi^2(2)$ then $P(X > 3.1) =$ ___0.212___.

$$P(X > 3.1) = e^{-3.1/2} = 0.212248$$

(b) **Using Chi-squared tables** determine the following: (use all decimal places available from the table)

(i) If $X \backsim \chi^2(18)$ then $P(X > 27)$ lies between ___0.05___ and ___0.1___. (You must use values from the Chi-squared tables.)

$$\begin{aligned} P(X > 25.989) &= 1 - 0.9 = 0.1 \\ P(X > 28.869) &= 1 - 0.95 = 0.05 \end{aligned}$$

(ii) If $X \backsim \chi^2(9)$ then the value of a such that $P(X \leq a) = 0.05$ is $a =$ ___3.325___.

(iii) If $X \backsim \chi^2(25)$ then the value of b such that $P(X > b) = 0.025$ is $b =$ ___40.646___.

(c) For the following questions specify **the distribution and its parameter(s)**:

(i) If $X \backsim G(-3, 2)$, $Y \backsim N(4, 16)$ and $V \backsim \mathit{Exponential}(2)$ independently then the distribution of

$W = \left(\frac{X+3}{2}\right)^2 + \left(\frac{Y-4}{4}\right)^2 + V$ is ___$\chi^2(4)$___.

$$\chi^2(1 + 1 + 2)$$

(ii) If $X_i \backsim \chi^2(i^2)$, $i = 1, 2, 3$ independently then the distribution of $\sum_{i=1}^{3} X_i$ is ___$\chi^2(14)$___.

$$\sum_{i=1}^{3} i^2 = 1 + 4 + 9 = 14$$