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 FALSE?

A: An estimator $\tilde{\theta}$ of a parameter θ is a random variable and its distribution is called the sampling distribution.

$\boxed{\text{B:}}$ The true value of the parameter θ is known as soon as we have collected the data.

C: An estimator $\tilde{\theta}$ of a parameter θ is a rule that tells us how to process the data to obtain a numerical value.

D: An estimate $\hat{\theta}$ is a known quantity which can be used to estimate θ.

(c) A function of the random variable $\mathbf{Y}$ and the parameter θ whose distribution does not depend on θ is called

A: an estimator for θ.

B: an estimate for θ.

C: a statistic for θ.

$\boxed{\text{D:}}$ a pivotal quantity for θ.

(d) The width of a 15% likelihood interval for θ based on data $y_1, y_2, \ldots, y_n$ from a $Poisson\,(\theta)$ distribution

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

B: increases as n increases.

C: decreases as n decreases.

D: does not change as n increases.

(e) The width of a 99% confidence interval for μ for data from a $G(\mu, \sigma)$ distribution with known value of σ and a fixed sample of size n

A: does not change as σ increases.

B: decreases as σ increases.

$\boxed{\text{C:}}$ increases as σ increases.

D: increases as σ decreases.

(f) In a Binomial(n, θ) experiment y successes in n trials were observed. Let thetahat $= y/n$. Which of the following functions in R should be used to calculate the relative likelihood function for θ?

A: BinRLF <- function(x) {y*log(x/thetahat)+(n-y)*log((1-x)/(1-thetahat))}

$\boxed{\text{B:}}$ BinRLF <- function(x) {exp(y*log(x/thetahat)+(n-y)*log((1-x)/(1-thetahat)))}

C: BinRLF <- function(x) {y*log(thetahat/x)+(n-y)*log((1-thetahat)/(1-x))}

D: BinRLF <- function(x) {exp(y*log(thetahat/x)+(n-y)*log((1-thetahat)/(1-x)))}

(g) In a Binomial experiment 28 successes in 50 trials were observed. Which of the following statements in R provides the maximum likelihood estimate of the probability of observing less than 14 Successes in 25 trials?

A: dbinom(14,25,28/50)

B: pbinom(14,25,28/50)

C: qbinom(14,25,28/50)

$\boxed{\text{D:}}$ pbinom(13,25,28/50)

E: qbinom(13,25,28/50)

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

$$f(y;\theta) = \frac{\theta}{y^{\theta+1}}, \qquad y \geq 1, \;\; \theta > 0.$$

(*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

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

or more simply

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

The log likelihood is

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

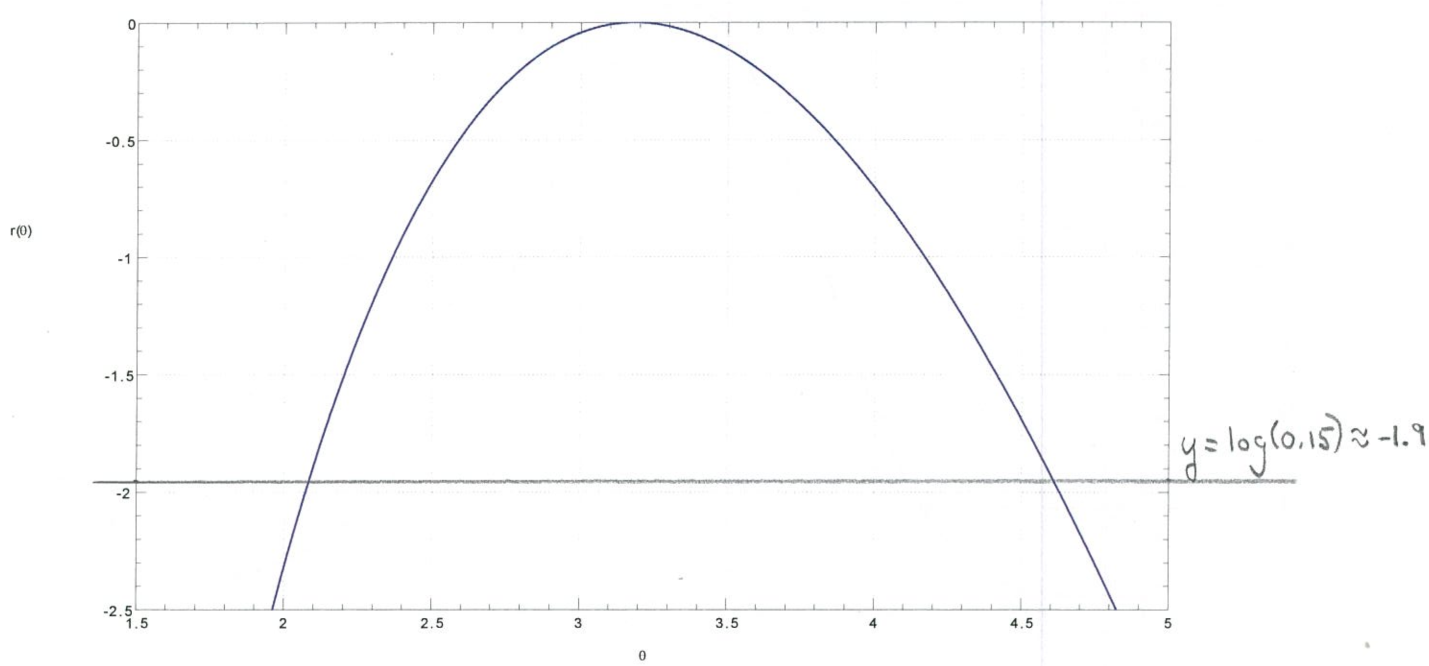
Solving

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

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 log relative likelihood function for θ for a given set of data:



From the above plot, graphically determine a 15% likelihood interval for θ (use one decimal place).

[2.1 , 4.6]

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

```
LogRLF <- function(x) {n*log(x/thetahat)+n*(1-x/thetahat)}
uniroot(function(x) LogRLF(x)-log(0.15),lower= 1.8 ,upper= 2.2 )
```

Note: Lower limit must be a number ≥ 1.5 but < 2.1. The upper limit must be a number > 2.1 but < 3.

3. [3 marks] In a study on the body temperatures of females after intense excercise, a researcher records the body temperature of 30 females aged $18-21$ after 20 minutes of intense exercise. Body tempertures in degrees Celsius are assumed to follow a $G(\mu, \sigma)$ distribution. A previous study has shown that $\sigma = 0.4$ degrees Celsius. Let y_i = body temperature of the ith female after 20 minutes of intense exercise, $i = 1, 2, \ldots, 30$. The researcher's data gave $\sum_{i=1}^{30} y_i = 1103.65$ and $\sum_{i=1}^{30} y_i^2 = 40608.95$.

Write your final answer ONLY in the space provided.

(a) [2 marks] A 95% confidence interval for μ is (use 3 decimal places)

[36.645 , 36.931]

$$\bar{y} \pm 1.96 \times \frac{\sigma}{\sqrt{n}} = \frac{1103.65}{30} \pm 1.96 \left(\frac{0.4}{\sqrt{30}} \right) = 37.78833 \pm 0.1431382 = [36.6452, 37.93146]$$

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

A. The probability that the true value of μ is contained in the interval is 95%.

$\boxed{\text{B:}}$ If the experiment is repeated many times, then we expect 95% of the intervals to contain the true value of μ.

C. We are 95% confident that $\mu = \hat{\mu}$.

D. All of the above.

4. [$7 \times 1 = 7$ marks] **For questions (a)-(c), 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 $P(X \geq 1.21) =$ ___0.271___.

$$P(X \geq 1.21) = 2P\left(Z \geq \sqrt{1.21}\right) = 2\left[1 - P(Z \leq 1.1)\right] = 2(1 - 0.86433) = 0.27134$$

(ii) If $X \backsim \chi^2(2)$ then $P(X > 1.5) =$ ___0.472___.

$$P(X > 1.5) = e^{-1.5/2} = 0.472367$$

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

(i) If $X \backsim \chi^2(7)$ then $P(X > 11)$ lies between ___0.1___ and ___0.2___. (You must use values from the Chi-squared tables.)

$$\begin{aligned} P(X > 9.803) &= 1 - 0.8 = 0.2 \\ P(X > 12.017) &= 1 - 0.9 = 0.1 \end{aligned}$$

(ii) If $X \backsim \chi^2(25)$ then the value of a such that $P(X \leq a) = 0.05$ is $a =$ ___14.611___.

(iii) If $X \backsim \chi^2(14)$ then the value of b such that $P(X > b) = 0.025$ is $b =$ ___26.119___.

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

(i) If $X \backsim G(1,2)$, $Y \backsim N(-1,9)$ and $V \backsim Exponential(2)$ independently then the distribution of

$W = \left(\frac{X-1}{2}\right)^2 + \left(\frac{Y+1}{3}\right)^2 + V$ is ___$\chi^2(4)$___.

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

(ii) If $X_i \backsim \chi^2(2i)$, $i = 1, 2, \ldots, 10$ independently then the distribution of $\sum_{i=1}^{10} X_i$ is ___$\chi^2(110)$___.

$$\sum_{i=1}^{10} i = 2 \times \frac{10(11)}{2} = 110$$