Statistics 2: Computer Practical Resit

1 Gamma random variables

If X is a Gamma(a, b) random variable, its probability density function is

$$f(x; a, b) = \begin{cases} \frac{b^a}{\Gamma(a)} x^{a-1} \exp(-bx) & x \geq 0, \\ 0 & \text{otherwise.} \end{cases}$$

The mean of X is a/b and the variance of X is a/b^2.

The parameter $a > 0$ is the "shape" parameter, while $b > 0$ is the "rate" parameter.

2 Estimation

Load the data in the file `data.Rdata`, which is a vector of $n = 100$ numbers stored in a variable called `data`.

```
load("data.Rdata")
```

It is decided that one should model the data as observed realizations of independent and identically distributed Gamma(a, b) random variables, $X_1, \ldots, X_n$.

Question 1. [2 marks] Derive the expressions for the method of moments estimates for a and b. Compute and report the estimates of a and b for the given data.

Question 2. [2 marks] Use the `optim` function to calculate numerically and report the maximum likelihood estimate of $\theta = (a, b)$ for the given data.

HINT: I suggest that you use the following function to compute the log-likelihood.

```
ll.gamma <- function(theta, xs) {
  return(sum(dgamma(xs, shape=theta[1], rate=theta[2], log=TRUE)))
}
```

Question 3. [2 marks] Approximate the mean-squared error of the Method of Moments and Maximum Likelihood estimators of each component of θ for $\theta = (a, b) = (2, 1.5)$ when $n = 100$.

Question 4. [2 marks] Derive the Maximum Likelihood estimator for b when a is known, i.e. only b is a statistical parameter.

Question 5. [3 marks]

a. Compute an observed 95% Wald confidence interval for b given the data, assuming it is known that $a = 2$.

b. Compute an approximation of the coverage of the corresponding confidence interval when $(a, b) = (2, 1.5)$ and $n = 100$.

3 Testing

Question 6. [3 marks] Derive the generalized likelihood ratio (GLR) test statistic for the null hypothesis that $a = 2$ (with no restriction on b), and state its asymptotic distribution.

Question 7 [2 marks] Report and interpret the p-value for the hypothesis test corresponding to Question 6, for the data provided.

Question 8 [2 marks] Consider a significance level of $\alpha = 0.05$, and assume that $b = 1.5$ and $n = 100$. Compute an approximation of the Type I error rate of the hypothesis test described in Question 6, and explain why it is not exactly equal to α.

Question 9 [2 marks] Consider a significance level of $\alpha = 0.05$, and assume that $b = 1.5$ and $n = 100$. Approximate the power of the hypothesis test described in Question 6 when $a = 2.5$, and explain the source of any approximation error.