ST227 - Exercise 3

Question 1

Assume the data

0.28031898	0.15079884	0.29919521	0.07400192	0.07468448
0.14209951	0.34165444	0.59901334	0.01924731	0.06012888
0.42266134	0.08853374	0.42016929	0.44196311	0.68137600
0.04048562	0.24948412	0.55145427	0.22625479	0.01550574

comes from an exponential distribution with rate parameter λ.

a. Derive the Method of Moments estimator for λ und compute it in R

b. In R, using `optim` or otherwise obtain maximum likelihood estimates for λ using the estimate from 1.a as starting values

c. Compare the estimates from 1.a and 1.b.

Question 2

This questions is divided into parts. Both parts use the same data set of fully observed lifetimes given below:

64	75	29	45	67	65	77	90	65	55
80	67	72	46	64	28	68	75	49	94

Let us suppose that this data set comes from a gamma distribution with shape-rate parametrisation, i.e:

$$f(x \mid \alpha, \beta) = \frac{\beta^\alpha}{\Gamma(\alpha)} x^{\alpha-1} \exp\{-\beta x\}, x > 0$$

a. Using the results:

$$\mathrm{E}(X) = \frac{\alpha}{\beta}, \quad \mathrm{var}(X) = \frac{\alpha}{\beta^2},$$

derive the method of moment estimators for α and β.

b. Using your MMEs above as the initial values for optim, derive the MLE for α and β.

c. Derive algebraically the probability density function for lifetime and write down the joint likelihood of the given sample.

d. Using the `optim` method in R, numerically obtain the maximum likelihood estimators of the model parameters.

Question 3

This question is divided into two parts. Both parts use the same data set of fully observed lifetimes given below:

80	75	38	45	62	65	77	92	65	60
55	67	72	46	64	35	68	52	45	94

Let us suppose that this data set comes from a Log-Normal distribution, i.e.,

$$f(x \mid \mu, \sigma) = \frac{1}{x\sigma\sqrt{2\pi}} \exp\left(-\frac{(\ln(x)-\mu)^2}{2\sigma^2}\right)$$

Using the results:

$$\mathrm{E}(X) = \exp\left(\mu + \frac{\sigma^2}{2}\right), \quad \mathrm{var}(X) = \left(\exp(\sigma^2) - 1\right)\left(\exp\left(\mu + \frac{1}{2}\sigma^2\right)\right)^2,$$

a. Derive the method of moment estimators for μ and σ^2.

b. Using your MMEs above as the initial values for `optim` or otherwise, derive the MLE for μ and σ^2, as the maximiser of the log-likelihood function.

c. Derive algebraically the expressions for the MLEs of μ, σ^2 and compare it to your solution from part b.

We propose a lifetime model with the following mortality intensity function:

$$\mu(t) = \lambda\gamma(\lambda t)^{\gamma-1}, \quad t \geq 0.$$

d. Derive algebraically the probability density function for lifetime and write down the joint-likelihood of the given sample.

e. Using `optim` and the initial values $\lambda_0 = 67$ and $\gamma_0 = 0.2233$, numerically obtain the maximum likelihood estimators of the model parameters.

Question 4 (Non-examinable)

The log-likelihood of a logistic regression model is given by

$$\ell(\beta) = \sum_{i=1}^{n} y_i\eta_i(\beta) - \log(1 + \exp\{\eta_i(\beta)\}),$$

where

$$\eta_i(\beta) = x_i^\top \beta = \sum_{i=1}^{p} x_{ij}\beta_j,$$

and where $\beta \in \Re^p$ is the model parameter, $y_1, \ldots, y_n \in \{0, 1\}$ are binary response variables and $x_1, \ldots, x_n \in \Re^p$ are p-dimensional covariate vectors.

a. In R, define a function `neg_loglikl` that takes as inputs
   1. `y`: A n-vector of response variables
   2. `X`: A $n \times p$ matrix of covariates (the ith row of X is the p-vector x_i in the equations above)

3. `beta`: The p-vector of model parameters and returns the negative of the log-likelihood of the logistic regression model.

b. Simulate data from a logistic regression model with $n = 1000$, $p = 10$ as follows:
   1. Construct β as from p i.i.d. draws from a standard normal distribution
   2. Construct X as a $n \times p$ matrix of i.i.d. draws from a standard normal distribution. Then, set the first column of X to all 1s.
   3. Draw a vector U of n uniform random variables. Compute the linear predictors $\eta = X\beta$ and the vector $\mu = 1/(1 + \exp\{-\eta\})$. Set $y_i = 1$ if $U_i < \mu_i$ and 0 else.

Set a seed for reproducibility.

c. Obtain the maximum likelihood estimate of β using the `optim` function. Use a p-vector of zeroes as initial values. Inspect the return of your optimisation, was optimisation successful?

d. For comparison, obtain the maximum likelihood estimates of β using the `glm` function. The syntax is `glm_model <- glm(y ~ -1 + X, family = binomial(link = "logit"))`. You can obtain estimates via `glm_model$coefficients`.